DOCUMENT RESUME

ED 240 164 TM 840 111

AUTHOR Power, Colin; And Others
TITLE National Assessment in Australia: An Evaluation of the Australian Studies in Student Performance Project.
INSTITUTION Australian Education Research and Development Committee, Canberra.
REPORT NO ERDC-R-35; ISBN-0-644-02012-1
PUB DATE 82
NOTE 157p.
AVAILABLE FROM Australian Government Publishing Service (AGPS), GPO Box 4, Canberra, ACT, 2601, Australia ($8.80).
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Educational Assessment; Expectation; *Foreign Countries; *National Programs; Program Effectiveness; *Program Evaluation; Sampling; Standards; *Testing Programs; Test Results
IDENTIFIERS Australia; *Australian Studies in Student Performance

ABSTRACT

An external evaluation of the impact of the Australian Studies in Student Performance (ASSP) Project on schools and systems is reported. The presented chapters follow in sequence the development of ASSP. The first chapter places the Australian Project in an international context and outlines the major questions dealt with in this evaluation. Chapter two traces the history of ASSP, while the next chapter maps expectations and opinions at the point of testing. In chapter four, technical issues relating to sampling and measurement of standards are addressed, while the next two chapters examine the administration and impact of the project in participating schools. Reactions to the results released in September 1981 are analyzed in chapter seven. The final chapter attempts to examine issues and decision possibilities with respect to national assessment in the light of the evidence collected. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED240164

National Assessment in Australia: An Evaluation of the Australian Studies in Student Performance Project

ERDC Report No. 35

Education Research and Development Committee

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

J. Ball

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Education Research and Development Committee

NATIONAL ASSESSMENT IN AUSTRALIA: AN EVALUATION OF THE AUSTRALIAN STUDIES IN STUDENT PERFORMANCE PROJECT

ERDC Report No. 35

Colin Power (Co-ordinator)	Flinders University of South Australia
Neil Baumgart	Macquarie University
Malcolm Eley	University of Tasmania
Mal Hewitson	Brisbane CAE, Kelvin Grove Campus
Donald Hogben	Flinders University of South Australia
Barry McGaw	Murdoch University
John Theobald	Monash University

Australian Government Publishing Service

Canberra 1982

© Commonwealth of Australia 1982

ISBN 0 644 02012 1

Printed by C. J. THOMPSON, Commonwealth Government Printer, Canberra

CONTENTS

		Page
Preface		v
Chapter One	Introduction	1
Chapter Two	Background to the testing program	8
Chapter Three	Expectations and opinions of ASSP	37
Chapter Four	The testing program	53
Chapter Five	Implementation of the testing program	70
Chapter Six	Impact of ASSP on participating schools	84
Chapter Seven	Reactions to the results of testing	95
Chapter Eight	Conclusions	116
Appendices		125
Index		148

PREFACE

In 1975 the Australian Council for Educational Research (ACER) undertook a survey of levels of literacy and numeracy in Australian Schools. Four years later, the Australian Education Council (AEC) commissioned the ACER to undertake a further study, called the Australian Studies in Student Performance (ASSP), on its behalf.

The findings of the 1975 ACER survey were given a great deal of publicity, much of it unfavourable to the education system. For example, they were used by those critical of the schools to support claims that standards in basic skill areas are not being maintained. The decision to undertake a further survey in 1980 met with resistance from the Directors General of Education, the Australian Teachers' Federation, the Australian Council of School Organizations and the Curriculum Development Centre. On the other hand, national assessment was supported by the Commonwealth, some State Governments, and employer organizations.

Given the controversial nature of the project, the Australian Education Council (AEC) at its June, 1980, meeting agreed to the recommendation that ASSP be "externally evaluated". It was intended that the evaluation should "examine, in particular, the impact of ASSP on schools and systems".

The present report stems from this recommendation. It was commissioned by the Educational Research and Development Committee (ERDC) following discussions with the AEC Steering group and representatives of national organizations likely to be interested in ASSP.

The design of the evaluation represents a co-operative venture. The conduct of the evaluation in each state was the responsibility of the team member from the state. Given the constraints of time and distance, individual members assumed primary responsibility for the development of instruments, data analysis and writing up of chapters or sections of the report, with the evaluation co-ordinator acting as editor. The chapters follow in sequence the development of ASSP. The first chapter places the Australian Project in an international context and outlines the major questions dealt with in this evaluation. Chapter Two traces the history of ASSP while the next chapter maps expectations and opinions at the point of testing. In Chapter Four, technical issues relating to sampling and measurement of standards are addressed, while the next two chapters examine the administration and impact of the project in participating schools. Reactions to the results released in September 1981 are analysed in Chapter Seven. The final chapter attempts to examine issues and decision possibilities with respect to national assessment in the light of the evidence collected. The team would like to acknowledge the assistance given to it by members of the AEC, its advisory committees, state and non-state systems and schools, the ACER and various interest groups who participated in the evaluation. Finally, we would like to give recognition to the financial support of the ERDC, and to express our gratitude for the advice and assistance given by its chairman (Mr. S. Dunn).

CHAPTER ONE

INTRODUCTION

Colin Power

Concern about standards of achievement and the introduction of mechanisms for monitoring the work of schools have been recurring themes throughout the history of education. In response to the recent resurgence of concern about standards, several governments have instituted national assessment programs designed to monitor the achievements of students. In that this report is concerned with the evaluation of one such program, it begins with an overview of developments in the USA, UK and Australia before proceeding to detail the problems to be addressed in the evaluation.

NATIONAL ASSESSMENT PROGRAMS IN USA AND UK

In the United States, national efforts to define and assess educational standards date back to 1964 when, as a result of criticisms of changes in school curricula and practice, public concern about standards of literacy and numeracy began to intensify. A National Assessment of Educational Progress (NAEP) project was begun in 1969. Since that time, the performance of USA students in a wide range of curriculum areas has been monitored. In the USA, there has been a dramatic increase in the extent of standardised testing, fostered, according to Resnick (1), by a widespread concern for rational management and the maintainance of standards in schools. At least 12 states have introduced "minimum competency" testing programs designed to monitor student achievement in "basic" skills areas; and minimum competency testing programs are being developed or considered by a further 26 states (2). For the most part, these programs were introduced by state legislatures and lay boards of education rather than professional educators - either as a reaction to what has been perceived as a public cry for accountability, or an urge to "legislate educational success without concern for methods or modes of achievement" (3).

Professional educators responsible for implementing testing programs initially tended to focus on the immediate problems of competency definition, test development, standard setting and program operation. Only recently has the debate turned to such larger issues as the need for competency testing, its effects on schools and its wider social and political ramifications. There is limited evidence (4) indicating that the results of national and state testing programs are largely ignored by policy makers, administrators and teachers; but whether such programs will ultimately lead to improved outcomes as hoped for by their advocates, or have the adverse effects predicted by their opponents, remains an open question.

In the U.K. the publication of the first of the Black Papers (5) demonstrated the growing concern of conservative groups over what they believe to be a shift away from traditional educational emphases and practices. In the speech which launched the so-called "Great Debate" in 1976, Labour Prime Minister James Callaghan stressed "the monitoring of standards", "the relationship between schools and industry", and the "need for a common curriculum". Central government initiatives aimed at pressuring schools to give greater attention to the skills and qualities valued by the government were already underway. In 1974, the establishment of a National Assessment of Performance Unit (APU), not unlike the NAEP in the United States, was announced by the Department of Education and Science. The APU is charged with the task of identifying, defining and testing, with a view to monitoring at a national level, standards of performance in the "major" curriculum areas. The areas in which testing is proceeding correspond closely to those which the government is now pushing as part of a common core curriculum for U.K. schools (6), viz. Maths, language (native and foreign) and science. However, Professor H. Goldstein of the University of London has raised serious questions about the validity of the statistical model with which the year by year comparisons of student performances are to be made (7). Others see the APU and the Department of Education core curriculum statement as symbolic of the growth in central intervention in, and bureaucratic control over, the business of schooling (8). In her analysis of the origins of APU, Broadfoot (9) suggests that national assessment is a late 20th century accountability mechanism which, like its 19th century counterparts, is designed to ensure that the state is getting value for the money it spends on education in the production of a literate and numerate workforce. In addition to the efforts of the central government to monitor standards at a national level, many local authorities have begun to use standardised tests which were designed originally for screening or selection purposes, to monitor standards and to allocate resources (10). There is, however, a growing disquiet about using the one program for many, and sometimes incompatible, purposes.

NATIONAL ASSESSMENT IN AUSTRALIA

Under the Australian constitution, the six State governments are legally and administratively responsible for education. To facilitate consultation and co-operation among governments, the Australian Education Council (AEC) was established in 1936. Its membership is restricted to the Education Ministers of the States, the Northern Territory and the Commonwealth. For the reasons elaborated in the next Chapter, the AEC decided in October, 1979, to commission the Australian Council for Educational Research (ACER) to undertake regular surveys of literacy and numeracy on its behalf. The basic aims, methodology and assumptions underlying the Australian national assessment project, now known as Australian Studies in Student Performance (ASSP), are similar to those of APU and NAEP. According to the AEC, the primary aim of ASSP is:

> to gather information on achievement in aspects of literacy and numeracy of a sample of 10 and 14 year old students throughout

Australia (11).

The project involved the development and administration of specially - constructed tests of basic reading, writing and numeration skills to "light" samples comprising approximately 1000 ten year old and 1000 fourteen year old students in each state, and 500 at both levels in the Australian Capital Territory and the Northern Territory. The tests were administered in October, 1980. The tests and sampling procedures used are similar to those employed in the 1975 ACER survey of literacy and numeracy (12). The design and background data collected mean that comparisons are possible at a national level by year (1975 vs. 1980), age (10 vs. 14 year old), sex (male vs. female) and location (urban vs. rural).

The tests developed by ACER for the ASSP project differ from conventional achievement tests. ASSP is designed to produce a general picture of basic skills achievement Australia-wide for the information of policy-makers and the public; conventional testing programs are designed to provide information about the achievements of individual students for use at the school or classroom level.

Compared with most other testing programs, the range of skills assessed by ASSP is very restricted. The test items sample only those minimum competencies of reading, writing and numeration which, it is believed, all students need to have mastered if they are to participate successfully in schooling and ultimately in the workforce and society in general. The tests employed can be described as "criterion-referenced tests". Judgments are made about performance with reference to a pre-specified criterion or "standard" rather than to the performance of others taking the tests. In the ASSP tests, a cutting score of 80 per cent is used; students answering 80 per cent of items correctly in a given test or subtest are certified as having "mastered" the skills being assessed. Issues relating to the measurement strategy used are discussed in more detail in Chapters Three and Four.

Through national testing, the AEC envisages making regular comparisons of student performance as a means of systems monitoring. It is anticipated that the results will be used to direct the attention of education systems and schools to basic skill areas which are considered to be in need of improvement. In addition to being used as a quality-control mechanism, ASSP is also likely to serve a public accountability function. Ministers and the community generally will have access to information about student achievement from which inferences will be made, rightly or wrongly, about the returns from public investment in education. The extent to which such functions can be served and their wider consequences are examined in subsequent chapters.

THE EVALUATION PROJECT

Concern over the possible backwash effects of the ASSP testing program on the curriculum led to a suggestion by the chairman of ERDC (Mr. S. Dunn) that there should be an "independent evaluative study" of the project. It was intended that the evaluation would look

particularly at the impact of ASSP on curriculum and teaching in schools and that the study would contribute to ASSP policy development and review. In June 1980, the AEC agreed in principle to the proposed evaluation. Following the AEC decision, individuals from each State with particular expertise in evaluation were invited to join the evaluation team. Team members discussed the broad scope of the evaluation with State interest groups and with the AEC steering committee for the ASSP project. Subsequently in July 1980, representatives of the major national interest groups involved in, or affected by, ASSP met with members of the evaluation team to identify the issues, questions and problems to be addressed by the evaluation. On the following day the team developed a framework for the evaluation study, taking into account the views and information needs of its audiences. The basic questions which the evaluation seeks to answer are as follows:

(i) How and why did Australia decide to initiate a national testing program?

(ii) What views are held within the community regarding the purposes and desirability of the program?

(iii) What are the strengths and limitations of the procedures used to assess standards of literacy and numeracy?

(iv) What problems, if any, were encountered in implementing the program?

(v) What impact did the testing program have, in the short-term, on teaching methods, curriculum, students, teachers and parents in participating schools?

(vi) In what ways were the findings interpreted and used by schools, education systems, policy makers and the media?

(vii) How did students, teachers, parents, interest groups and the media react to the results?

(viii) What are the likely long-term effects of ASSP on students, schools and society?

In reading any evaluation report it is important to be aware of the framework of assumptions and values underlying the evaluation and the program being evaluated. National assessment programs represent the outcome of analyses of the presumed shortcomings of the education system in relation to the needs of society from what Broadfoot (14) describes as a "technological" framework. Assumptions underlying this framework include the following:

(i) Consensus exists as to the skills and standards which need to be achieved if students are to participate in further education and society;

(ii) All children in regular schools can and should achieve at least the same minimum standards of literacy and numeracy;

(iii) Governments, the public and educators need to know what standards are being achieved by students throughout the country;

(iv) School systems should be accountable to governments and the public for ensuring that the minimum standards expected are achieved; and

(v) The technology exists or can be created by means of which standards can be monitored.

Advocates of national assessment further believe that regular nation-wide assessments of basic skills using criterion-referenced tests and a "light" sampling approach is a feasible answer to the problem of providing the information and controls needed, without having serious backwash effects on schools.

By and large, the evaluation has been undertaken within the parameters of the prevailing framework. We have accepted that schools should be held accountable for their work; that the achievements of students need to be assessed; that improvements need to be made in the teaching of reading, writing and mathematics. We have proceeded to address the questions being asked by the sponsors and developers of the project (AEC and ACER) about its technical soundness, implementation and impact on schools. However, any evaluation of ASSP cannot avoid the questions being asked about the desirability of national assessment and its broader consequences for students, schools and society.

In the concluding sections of this report we have attempted to examine the goals and consequences of ASSP from a number of different perspectives, and to clarify what we believe to be the major issues which still need to be faced in reaching decisions about the future of ASSP.

REFERENCES

1. Resnick, L.B. Introduction: research to inform a debate. Phi Delta Kappan, May 1981, 623.

2. Jaeger, R.M. & Kehr, C. Minimum competency achievement testing. Berkeley, Cal.: McCutchan, 1980.

3. loc.cit. p.vii.

4. Resnick op. cit.

5. Cox, C.B. & Dyson, A.E. (eds.). Fight for education (Black paper). London: The Critical Quarterly Society, 1969.

6. Department of Education and Science. The school curriculum. London: HMSO, June 1981.

7. Goldstein, H. Consequences of using the Rasch model for educational assessment. British Journal of Educational Research, 1979, 5(2), 211-220.

Goldstein, A. & Blinkhorn, S. Monitoring educational standards -an inappropriate model. Bulletin of the British Psychological Society, 1977, 30, 309-311.

8. Elliott, J. The Cambridge Accountability Project, Cambridge Journal of Education, 1981, 11(2), 146-166.

9. Broadfoot, P. Assessment, schools and society. London: Methuen, 1979.

10. Blackstone, T., Goldstein, H., Wood, R. & Gipps, C. Evaluation of testing in schools project: Report on the survey of testing activities in local authorities. London: Institute of Education, University of London, June, 1981.

11. AEC Press release, July, 1980.

12. Bourke, S.F. & Keeves, J.P. Australian Studies in School Performance. Vol.III The Mastery of Literacy and Numeracy. Final Report. ERDC Report No. 13. Canberra: AGPS, 1977.

13. NAEP and APU cover a wider range of curriculum areas.

14. Broadfoot op. cit. p.131.

CHAPTER TWO

BACKGROUND TO THE TESTING PROGRAM

Colin Power

INTRODUCTION

According to Schon(1), there are thresholds of change in education - critical levels of energy which must be reached to precipitate a change of state. In this chapter, it is argued that the development of ASSP must be understood in relation to the political and ideological change of state precipitated by the economic crisis of the mid 1970s. This crisis led to high levels of unemployment amongst school leavers and to the reemergence of standards of student performance as a politically salient issue. The debate which has followed has been largely ideological. It has led to a number of federal initiatives (including national assessment) which reflect the politics of austerity and the philosophy of neo-conservatism, the political and economic tide having turned against the liberal and developmental perspectives encouraged in the early 1970s.

The ideological context within which ASSP developed includes a reassertion of utilitarian priorities and values in education; the conviction that there should be a closer relation between education and industry; and a general scepticism about the cost-effectiveness of the public sector. Ministerial responsibility, systems accountability for outcomes, the lack of "hard data" on standards, and the need for quality-control mechanisms with adequate safeguards have been key issues in the development of the program. Finally, it is argued that the policy making process involves the formation of coalitions, negotiation, consensus and conflict. Hence, the particular form of national assessment which has evolved in Australia reflects the respective powers, interests, values and strategies of the Ministers, Directors-General, professional bodies and pressure groups involved.

GENESIS OF ASSP

Initially, education in Australia was a private enterprise. The intrusion of the Australian colonial governments in the latter part of the nineteenth century was the result of a build-up of concerns about social order, the assumption being that a literate and numerate populace directed along the proper path was essential if Australia was to develop into a stable, cohesive and prosperous nation. The expected emphasis in State school systems thereafter has been on developing efficient means of accomplishing this task.

The post World War II decades saw great changes in Australian society. Its rapid growth in size resulted from a vast influx of migrants and a rise in birth rates. Economic development matched the growth in population. With increasing urbanisation, secondary and tertiary industries employed a much greater share of the workforce at the continuing expense of the rural sector. Under these conditions, there were demands for more and more education as individuals and governments began to view education as a form of investment. Initially, the physical problems of providing schools, teachers and equipment dominated the concerns of parents, pressure groups and politicians. While there were some changes in the structure and content of education in the 1960's a threshold of change was not achieved until the early 1970's when the Labour government, capturing the spirit of the time, came to power. Education was the cornerstone of Labour's vision for a just and equitable society. It moved quickly to establish the Australian Schools Commission. Not only did the government through the Commission assist in meeting the material needs of schools, but the Commission fostered innovation in school programs and practices, devolution of responsibility, community participation and concern over the educational needs of disadvantaged groups in our society.

The concern for children with learning difficulties which might impede the acquisition of basic skills necessary for effective participation in a technological society led to the establishment of the House of Representatives Select Committee on Specific Learning Difficulties on 31st October, 1974. The Committee took the view that society has an obligation to its weakest members and would not wish them to fall below certain minimal standards of achievement. While the Committee felt it would be undesirable to define these standards with precise measures of literacy and numeracy, it suggested that levels of achievement are acceptable "only when they enable the individual to compete in society in such a way that he can, with reasonable effort, obtain for himself reasonable livelihood, dignity and reasonable social intercourse" (2).

The first term of reference of this Committee was to inquire into the incidence of all forms of specific learning difficulties among Australian children and adults. The Committee found that it could not reliably determine the magnitude and extent of the difficulties in the Australian population on the basis of the available evidence. Thus in February 1975, the Chairman of the Committee (at that stage Mr. C.R.T. Mathews), approached the ACER for assistance. ACER suggested that a national survey of learning difficulties among Australian school children be undertaken. According to Dr. Keeves (the present Director of ACER), Mr. Mathews hoped that there would be sufficient changes in educational policy and practice flowing from the Inquiry to make it worthwhile to undertake further studies aimed at determining if improvements had taken place. The proposal for the project was referred to the AARCDE (now ERDC) so that the survey could be adequately financed.

The ACER was asked to provide by the end of 1975 evidence of performance on the basic skills of reading, writing and numeration, and estimates of children failing to attain these skills. The ACER

study involved the construction of new types of instruments (called criterion-referenced tests) and the testing of almost 13,000 ten and fourteen year old students throughout Australia. The ACER was under great pressure to acquire suitable staff, prepare the tests and organize the testing on time; despite the difficulties, the testing went ahead as planned in October 1975.

Early in 1976 interim reports of the ACER project were forwarded to AARCDE, the Minister (Senator Carrick) and the Select Committee. Subsequently reports of the project were published by the ERDC (3).

The ACER findings featured prominently in the report of the Select Committee. The Committee took the view that the ACER survey confirmed that there was a problem in Australian schools and that a significant number of children were failing to reach "adequate" levels of literacy and numeracy. The Committee also concluded that it was not possible to estimate whether standards of literacy, numeracy, reading and spelling were declining, improving or remaining stable. In its report tabled in parliament in October, 1976, the Committee supported the notion of national monitoring of basic skills:

> The incidence of learning difficulties among Australian school children should be determined at regular intervals by a nation-wide survey based on similar principles to the ACER study 'Literacy and numeracy in Australian schools' (4).

THE STANDARDS DEBATE

At the time of the release of the ACER results, Australia was once more on the threshold of change. The 1960's and early 1970's represented an unprecedented period of economic growth, full employment and faith in the capacity of education to help create the conditions for a prosperous and equitable society. Commonwealth and State governments had increased their expenditure on education and allowed schools greater freedom to innovate and to modify their programs to cater for the diversity of needs and interests of students. The mid-1970's brought with them a sudden downturn in the national economy, high rates of youth unemployment, demographic changes and pressure to reduce expenditure on education. The conditions prevailing in Australia precipitated a resurgence of conservatism and change in government.

In Australia, as in the United States (5) and Britain (6), the issue of standards became a central concern in public debates about education. Employers and tertiary educators had become very critical of the level of academic achievement of school leavers (7). In their view, standards had declined in recent years because too much emphasis had been given to "progressive" ideas at the expense of "basic" knowledge and skills (8). Pressure groups were particularly concerned about what was seen to be the erosion of intellectual, social and moral values in schools. Reactionary groups like the Committee against Regressive Education, the Campaign for Responsible Education and the Committee to Restore Educational Standards have sought to return education to the tried and tested practices of the past and to the "three R's".

The Australian Council for Educational Standards (ACES) has been concerned more with strengthening the intellectual core of education than "back to the basics". To this end, the ACES has sought by "vigorous public discussion to resist attempts to impose an anti-intellectual approach in the name of progressive education" (9). While ACES has been largely concerned with promoting intellectual inquiry through the academic disciplines, it is concerned about standards of literacy and numeracy for educational rather than economic or social reasons, since a sound grounding in "basic knowledge" is a prerequisite for secondary and tertiary studies. A key figure in ACES has been Leonie Kramer, Professor of Australian Literature at the University of Sydney. She has been highly critical of the emphasis on "personal development" and the "lack of rigour and objectivity" in school curricula, and the "decay of the disciplined study" of English, History and Geography in secondary schools.

At the very time at which the conservative backlash was building up to that crucial level of energy needed to precipitate a change of state, a draft of the ACER study was "leaked" to the Bulletin. In a biting attack on Australia's education system, Peter Samuel (10) stated that the results of the ACER survey revealed a decline in the quality of Australian education alleged to be the "most alarming we have ever published". The article went on to suggest that while the results were not uniformly depressing, they did "document a massive failure of education in primary and secondary schools". The Bulletin called for its readers to answer such questions as "What is wrong with education?" and "Why are we paying more and educating less?" The attack was taken up by other sections of the media. The Australian Teachers Federation, the Chairman of the Australian Schools Commission (11) and some Councils of Parents and Citizens' Associations mounted a defence, arguing that there was no evidence of a decline in standards and that the origins of the problems of some school leavers were complex and could not be attributed solely to schools. Dr. W. Radford, Director of ACER, wrote to the Bulletin objecting to the misrepresentations of the ACER results in the Samuel's article. His efforts to explain the meaning of the data to the public and his entirely proper objections were given a particularly inept and ungracious treatment. Samuel dismissed the objections asserting that just because there is a lack of evidence of a deterioration in education "does not mean that there has not been such a deterioration".

The mass media were instrumental in exploiting the concerns about education by presenting the results of the first national survey of literacy and numeracy in a sensational and biased way. The ACER in its report of the project commented:

> In the debate that has followed, many statements have been made which have misinterpreted and misrepresented the actual results of the investigation (12).

The treatment which schools received in the media over the results of the 1975 national survey created resentment amongst the senior officials of education systems, teachers' unions and state school parents' organizations, most of whom are suspicious of the uses

to which the results of any system of national assessment might be put.

At the Commonwealth level, the debate over standards and the issue of national assessment mirrored the public debate. The political debate was conducted at a time of economic crisis and in a context within which a radical re-ordering of financial priorities in harmony with Liberal philosophy (13) was being undertaken.

Senator P. Rae (14) asked for statistics going back twenty-five years (with a breakdown by State and Territory) on students known to reach secondary school without being able to read and write, as well as causes of this. Senator J. Carrick circulated a reply (15) stating that the required statistics were not available. He indicated that baseline data were, however, being gathered by means of the 1975 ACER survey being undertaken for the House of Representatives Select Committee on Specific Learning Difficulties. This survey would produce reliable national data for the first time, according to Senator Carrick, and would provide a benchmark for measuring changes in standards over time. From this time onwards, it became clear that Senator Carrick and the government were convinced that a national monitoring system was needed.

Senator Wood (16) put to Senator Carrick a question without notice regarding a press report of students "uneducated to the point of being illiterate" and absolutely ignorant of geography (not knowing where New York or Fiji are). In view of money spent on education, he asked whether the Minister and the Government will make it mandatory that children are taught these basic elements of education. While agreeing with the seriousness of the matter, yet recalling that "as far back as the classics we find that Socrates and Diogenes were writing about and regretting the growing illiteracy of the younger generation", Senator Carrick stated that the ACER report was due shortly. He also expressed concern about the exclusive pursuit by some of innovation and experimentation, considering it necessary first to ensure basic skills, and then to proceed with innovation and experimentation.

Senator Carrick (17) presented a Report on Literacy and Numeracy in Australian Schools to the Senate on 25th May, 1976, and asserted the need for regular testing with appropriate time intervals in between, if answers are to be given to questions about changes. He indicated that he had asked the ERDC to examine the complex problems involved in a program of national assessment. The need for regular testing was reiterated when Senator Carrick answered a question from Senator Davidson (18) regarding a press report of a statement by Dr. McKinnon of the Schools Commission about the resistance of the Australian school system to reform. Again Senator Carrick declared his belief in basic skills as vital to the development of other educational skills. The same message was repeated the following day to Senator Cameron: without the basic skills of literacy and numeracy, no kind of innovation in education can be successfully embarked upon (20).

Speaking to his report in October 1976, Mr. Cadman (19) emphasised that "the classroom teacher must be the key in any attempt to overcome learning difficulties" and hence there is need for more adequate pre and in-service training of teachers than has been provided hitherto. In order to provide information about the nature and incidence of numeracy and literacy problems, it was strongly recommended that regular national surveys, similar to the 1975 Survey, be conducted.

The need for national testing at regular intervals was endorsed by Mr. K. Beazley (20) (former Labour Minister for Education who authorized the 1975 investigation into literacy and numeracy in Australian Schools) and Mr. Hyde during debate on the Report on Learning Difficulties in December, 1976. Both speakers referred also to the Fitzgerald Report, Poverty and Education in Australia, issued to them that day, in the contexts of resource allocation and teacher competency. Mr. Hyde specifically mentioned minimum standards pointing out the difficulty of competing in today's society "if you cannot learn simple arithmetic, simple reading, simple literacy and simple numeracy". The uselessness of some aspects of teacher education programs was the subject of Mr. Sainsbury's (21) speech in reference to the Report on Literacy and Numeracy. Cooperation between the States and the Commonwealth on teacher education was another matter raised in debates on both the Literacy and Numeracy and Learning Difficulties Reports.

In reply to an enquiry by Senator Rae on what was to be done about the Report on Learning Difficulties, Senator Carrick (22) stated that steps must be taken to enable teachers to recognise learning problems in the classroom, and that all teachers should be remedial teachers. To ensure that such steps were taken, he indicated that in the previous week (February, 1977), the AEC had decided there should be a National Inquiry into Teacher Education.

On successive days in November 1978, Senator Carrick (23) referred to national testing, on one occasion stating that a study group was looking at the feasibility of such a program, and on the other expressing the need for it as evidence of whether schools are succeeding or failing in the teaching of basic skills. He frequently drew attention to the National Inquiry into Teacher Education, which would, according to its terms of reference, deal with upgrading basic skills as well as the effectiveness of training in early detection, diagnosis and remediation of learning difficulties and handicaps. Cooperation with the States in the development of testing measures to assist classroom teachers and systems to identify deficiencies in literacy and numeracy was reported in answer to a question arising from a press report of functionally illiterate people in Sydney (24). Senator Carrick said that the States were "taking their own action in an area which is their responsibility" in consultation with the Commonwealth.

One of the areas of great concern during the present economic crisis has been the relationship between the world of work and the world of education.

The Commonwealth government established a Committee of Inquiry into Education and Training (Chairman: Professor B.R. Williams) in 1976. Having reviewed the criticisms of schools systems and the evidence on student peformance in aspects of literacy and numeracy (25) the Williams committee concluded that complaints were "well-based". In its view, many school leavers were not sufficiently literate and numerate (26). The Committee was sympathetic to the case put forward by the ACER and Senator Carrick for a national assessment system. The Committee had some reservations about the adoption of a standardised national testing program as a response to criticisms of achievements in schools, but argued that it is important to monitor attainments in the basic skills once these have been specified. The report recommended that the Australian Education Council finance the ACER to undertake a study so that it will be possible:

(a) to specify minimum acceptable performance levels in basic skills for particular groups;

(b) to specify a range of performance levels to be expected of students of varying ability; and

(c) to develop appropriate testing devices to enable schools - and on occasion the nation - to check on achievement without imposing a straightjacket on schools.

Commenting on the report, the Australian (27) reiterated its previously expressed belief that parents were concerned, even anxious, because their children were leaving school lacking the fluency in basic skills needed to cope with modern life. The Australian reiterated criticisms of teacher education institutions for the lack of emphasis on methods of teaching of basic skills in their programs. State Education Departments were criticised for being "locked into the so-called 'progressive' approaches" and for refusing to take a "serious look at the problems". A most scathing attack on the performance of schools, C.A.E.'s and universities was issued by the Prime Minister, Mr. M. Fraser, in an address at the national convention of the Young Liberal Movement (28). Despite massively increased expenditure and lower class sizes, he claimed young people were being sent out of school unable to read, write or add to an acceptable standard. The education system, he said, was therefore to blame for much of our youth unemployment in Australia. Not surprisingly, the Prime Minister's claims were hotly denied by the Australian Teachers Federation. The Federation claimed that standards have not declined, and attributed the blame for youth unemployment to government economic policy.

It should be noted that criticisms of schools over standards, and demands that schools and teachers be held accountable for student performance are neither new nor unique. Historical analyses have shown the cyclic nature of demands for accountability and the part played by testing systems to maintain control over the content and quality of teaching (29). During times of economic crisis, schools seem to be particularly vulnerable: the costs of schooling and the standard of performance invariably are at issue. For instance, the conditions in the 1930's drew forth attacks on "progressive" ideas and

practices followed by demands from businessmen and politicians for a return to the "basics", a stiffer pedagogy, and closer links between education and work (30). This pattern of events repeats itself from time to time in every nation. Under favourable economic conditions, the controls over education are relaxed a little and schools are encouraged to adopt a less rigid and restricted approach to their task. Economic downturns often precipitate criticism, a reassertion of political control and cuts in education expenditure.

As Bessant and Spaul (31) point out, the institutions of schooling have always been too important to be left to their own devices for they are closely tied to the future of the state itself. The subservience of schools to the state in Australia was clearly apparent until the late 1960's. As our political institutions have matured, the relationship between school and state has become more subtle. But no matter how it is accomplished, the activities of schools are likely to continue to be constrained, governed, monitored and regulated by the state.

The standards debate has a strong ideological dimension. Substantively, conflict and argument have revolved around three issues - content, pedagogy and the control of education, and thus implicitly the purpose of education as it is conceived by egalitarian liberals has been called into question.

The general line of argument developed in the debate by the proponents of national assessment has been as follows:

(i) Many groups in the community are concerned about standards of literacy and numeracy in schools.

(ii) A strong grounding in literacy and numeracy is vital; individuals lacking basic skills are unlikely to make satisfactory educational progress or to gain employment on leaving school.

(iii) While there is no substantive evidence of decline or improvement, at the national level, the 1975 ACER survey confirmed that a significant proportion of students were leaving school lacking in basic skills.

(iv) As learning depends on teaching, the adequacy and appropriateness of the teacher's education is important; while a great deal of money has been spent on pre- and in-service teacher education, there is no evidence that this expenditure has been worthwhile, indeed it is suspected that if anything, less emphasis is being placed on the teaching of basic skills.

(v) In order that answers to questions about the effectiveness of policies, programs and innovations funded by governments can be answered, it is desirable educational standards be monitored.

THE ERDC STUDY GROUP

Prior to the tabling of the House of Representatives Select Committee Report in 1976, Senator Carrick had expressed his interest in the problem of educational standards and the desirability of monitoring changes in student performance. Following an examination of material describing the NAEP project in the United States provided by ERDC, the Minister, on 24th February 1976, wrote to the Chairman of ERDC (Mr. S.S. Dunn) "expressing support for the ACER study on literacy and numeracy and inviting his Committee to consider how best to develop a program for systematic evaluation of education standards" (32). Senator Carrick's letter stressed the need for "establishing benchmarks and monitoring change in all relevant skill areas which the community expects to be maintained and/or achieved through education".

The Minister's letter resulted in the establishment, in the latter half of 1976, of an ERDC Study Group "to consider the need for a national assessment program in Australia" (33). The Study Group was formed under the chairmanship of Dr. W.C. Radford, Director of A.C.E.R. On the death of Dr. Radford in November 1976, Mr. A.H. Webster (N.S.W. State Education Department) was appointed chairman.

As the chairman of the ERDC Study Group makes it clear in his preface to the final report, consensus on the issues of national assessment was reached only occasionally(34). Overall, the group was strongly opposed to a national assessment program as large and complex as NAEP in the United States. Members took the view that the individual school must be the focal point of action in any effort to improve the quality of education. The report therefore argued that the school should be the primary agent in carrying out assessment which has as its purposes the monitoring and improvement of the overall quality of education. The bulk of the Study Group recommendations relate to ways of increasing the competence of schools and teachers in assessment. At the same time, however, a majority of members of the Study Group acknowledged that systems administrators and/or Ministers may wish to conduct state or national surveys of basic skills since these

> ..provide a useful public evaluation of certain basic activities in Australian schools (35).

The Study Group cautioned that such surveys should not pretend to be a substitute for wider accountability. While some of the members (36) of the Study Group had serious misgivings about the desirability of any form of national assessment, the report eventually recommended that the Commonwealth Government should fund ACER to repeat the basic literacy and numeracy tests on a national basis. In an attempt to pre-empt the anticipated undesirable backwash effects on school curricula and misuses of the results, the Study Group suggested that a consultative committee be established to work with the ACER.

Various versions of the Study Group report and ultimately an interim report were circulated during 1977 and 1978 by the ERDC on a confidential basis to twenty-seven functionally involved organizations and individuals for comment, and to each Minister of Education for information. Responses to the draft and additional information on national assessment in other countries were taken into account by the

Study Group in producing the final report, the published version of which did not become available until 1980. The final report contains an ERDC Annexe in which ERDC indicates that there are good reasons for monitoring student performance at a state or national level, even though there are dangers. It argues that a national assessment program could be expected to yield "education indices" akin to productivity indices or unemployment rates, and that, if collected regularly, such indices provide measures of change. While indicies can act as a warning system, they do not, in themselves, provide information about causes or indicate suitable remedies. Having raised a number of issues relating to the scope, content and frequency of testing, the ERDC Annexe comes out in favour of a restricted basic skills monitoring program to be conducted by the ACER in consultation with an appropriate Steering Committee.

The reactions to the Report reflect its ambiguity. The Directors-General interpreted the report as being supportive of a school-based testing program (Recommendations 2 to 7) and opposed to national assessment (Recommendation 1). Senator Carrick and his supporters saw the report as legitimising their conviction that a national system for monitoring literacy and numeracy was needed (Recommendation 8).

BASIC SKILLS ASSESSMENT IN THE STATES

Under the Australian constitution, education is a state responsibility. Over the years, each state has developed its own system for assessing the performance of students and teachers. Until quite recently, state systems were highly centralised; schools and teachers were controlled by state prescribed syllabuses, public examinations and the inspectorial system. Pressure from teachers' organizations and social changes in the 1960's led to a weakening of these controls. To varying degrees, State schools have been given greater responsibility for their own teaching and evaluation programs. Despite this, the vast majority of Australian primary schools still regularly assess student reading and mathematics achievement, either as part of a state-wide testing program or using testing materials purchased from the ACER. The Report of the ERDC Study Group on National Assessment contains details of the diversity of basic skills testing programs, surveys and developmental projects (eg. NSW Primary Evaluation Program, Queensland Test Development Project, item banking projects in Tasmania and WA). It also indicates the range of testing materials available from ACER which are used by schools and systems.

The states have been quite willing to co-operate with each other in national projects aimed at developing diagnostic basic skills testing programs for use at the school level. However, the WA Education Department in particular, takes the view that the monitoring of student performance in key areas of the curriculum is a state, not a national, responsibility.

By and large, basic skills testing programs have been under the control of schools and State Departments. They determine who and what is assessed, and by whom and for what purposes the results are used. State-wide surveys of basic skills performance are usually designed to

meet the information needs of the professionals rather than those of the public. In some states, monitoring of student achievement in basic skill areas has been under way for a long period of time and the results have been published on a regular basis. For example, the Queensland Education Department (37) was able to produce evidence that standards of mathematics in that state are high in comparison with other systems, and that standards of reading have improved and were considerably higher in the 1970's than in previous decades. However, for the most part, testing programs in the states are designed to satisfy the purposes of the profession rather than those of governments. It is not possible to use the results to judge whether standards are improving or declining. It was this lack of public information about standards, and the reemergence of the view that schools and State Departments should be held accountable for the performance of their students, which led the then Victorian Minister of Education (Hon. L. Thompson) in 1977 to seek to introduce a massive basic skills testing program; the ill-fated Essential Skills Assessment Project or ESAP.

The aim was to introduce a "regular monitoring process" of the educational attainments of students in areas of "basic achievement". The areas to be tested were those considered "essential for worthwhile participation" in later schooling, employment and society. An Assessment Committee was established in October, 1977 by Ministerial invitation. The Committee was chaired by Professor K.B. Start of the University of Melbourne. Professor Start, an advocate of external standardised testing, was a member of the ERDC Study Group and had previously worked at the National Foundation for Educational Research in the U.K. on surveys of reading standards (38). While the three member committee included Dr. G.R. Maddocks, Assistant Director-General of Education (Curriculum and Planning), the Committee was directly responsible to the Victorian Minister. In effect, the Education Department and the Director-General were by-passed. The Minister asked the Committee to develop a program which would assess the literacy and numeracy of children in Victorian schools in such a way that comparisons could be made, from year to year, between major differences in curriculum emphasis and among major groupings in the population, for example, boys and girls, urban and rural.

It was intended that information on the literacy and numeracy of Grades 3, 6 and 9 children be fed back to schools and regions and used to identify areas of weakness in school programs. In the long run, it was hoped that standards of literacy and numeracy of all children, particularly those considered in need or at risk, would improve. It was also suggested that schools could use the information in their submissions to Division Directors seeking support for programs aimed at remedying deficiencies.

The teachers' unions were concerned that the results would be used as a club to beat teachers and the Department. Many teachers were concerned that ESAP would force schools to stress literacy and numeracy and to neglect other areas and objectives. The fear was that the end result would be a centralised curriculum which would not be in the children's best interests. The opposition to ESAP from the Executives of three of the four Victorian Teachers Unions was total

and drew into it the executive of parent groups. According to Professor Start (39), at over one hundred meetings held with groups of teachers throughout the state, the implacable opposition to external standardised testing was reiterated by representatives of these unions. He also claims that the State Department showed a reluctance to co-operate in the development of a program over which it had little control. On the other hand, it has been claimed that ESAP and Professor Start were overly ambitious and the sheer size, cost and complexity of the operation made it impractical. For whatever reasons, progress on the project was slow and negotiation with the unions was protracted and unproductive. The withdrawal of Professor Start due to serious illness in 1979 created both a hiatus and an opportunity for the new Minister of Education (Hon. A. Hunt) to abandon ESAP. The Victorian cabinet had decided that ASSP with its light sampling and confidentiality provisions should meet the basic objections to ESAP held by the unions. At the same time, it believed that ASSP would provide information on the "status of student performance in basic skills to replace unsubstantiated rhetoric" and aid in "formulating policy relating to the provision of resources" (40).

Of all the State Ministers, Thompson and, subsequently, Hunt and Lacy were strongest in their support of the Carrick proposal to establish a national system for monitoring standards.

THE ACER SCHOOL ACHIEVEMENT TESTING PROPOSAL

Concurrently with deliberations of the ERDC Study group, plans for the development of a school-based achievement testing program were being developed by the ACER. When the ACER had been asked in 1975 to survey student achievement in reading, writing and numeration, it was Dr. Keeves' hope that this was to be the first of several surveys. Shortly before his appointment to the Directorship of ACER in June, 1977, Dr. Keeves submitted to the Council plans for the future research efforts of the organization. His submission to the July Council meeting argued that ACER should develop a range of criterion-referenced tests which could be used for the assessment of student performance in schools. In October 1977, the Council agreed that ACER should proceed with the development of these tests, it being known that the ERDC Study Group and the Williams Committee were likely to come out in favour of the development of tests which could be used by classroom teachers, schools and/or systems to measure the progress of students. At the 1978 ACER Council meeting, Dr. J. Izard (Assistant Director, ACER) outlined a proposal for a school-focussed achievement testing program. It was envisaged that basic skills tests would be developed co-operatively by the states and ACER at two levels: review tests designed to enable schools to screen their students and programs for learning difficulties, and more detailed diagnostic progress tests designed to enable teachers to pinpoint the difficulties of students at risk. The proposal was supported by the Council and, at this stage, was seen by systems representatives on the Council to be a positive educational move. Later that year, representatives of the states, ERDC and ACER met to explore the feasibility of a co-operative venture along the lines proposed by Izard. This venture has since become known as the Australian

Cooperative Assessment Project, ACAP.

In February 1979, the ERDC Study Group Report was tabled and discussed by the AEC. The ACER was asked to prepare and submit a more detailed proposal for an achievement testing program. Whereas the states drew a sharp distinction between national survey testing and the school focussed Progress-Review testing, the proposal developed by Mr. S. Bourke (Chief Research Officer, ACER) later in the year, sought to tie the two together. The proposal, which was strongly supported by the Director of ACER (Dr. Keeves), envisaged that an interdependent set of tests would be developed at three levels:

1. 'Progress' tests for diagnostic use by teachers within classrooms;

2. 'Review' tests for use by teachers and principals in summarizing achievement in certain areas;

3. 'Survey' tests for use by those responsible for education systems.

The ACER recommended that, initially, work should begin on Progress and Review tests covering the curriculum areas of reading, writing, mathematics, listening and speaking skills, inquiry skills, social learning and natural science, suitable for use by schools in the last years of primary schooling. Plans for the national survey tests were to include only the areas of reading, writing and number. An in-service education program for teachers and principals to assist in the appropriate use of the Progress and Review tests, and in the interpretation of information arising from the Survey tests, was also recommended.

The total cost of the survey test program, recommended to run for at least five years, was estimated at $650,000. It was anticipated that a three-year development plan for the Progress and Review tests would cost approximately $400,000.

Once again reactions to the ACER proposal varied according to the particular interests and ideologies of Ministers and their advisors. Ministers who saw a need for a strengthening of conventional notions of Ministerial responsibility and public accountability concentrated on survey testing proposals, while Directors-General focussed on the proposal to develop Progress and Review tests. ACER's concern to ensure the proposal was accepted as a package deal reflected its interests as the major producer of testing materials in Australia.

THE ROLE OF THE AEC

AEC meetings have provided an opportunity for States to seek Commonwealth support for specific educational purposes and for successive Commonwealth governments to press their national education priorities. The Liberal-Country Party government which held power from 1949 to 1972 was averse to national planning, but gradually increased its involvement in providing support for education as it reacted to the lobbying of pressure groups and the increased demand for secondary and tertiary education. Its methodology has been described as "ad hoc intervention" with the government emerging as a

"paternalistic, developmental protection agency and a manager of crisis" (41). But basic problems relating to responsibilities and resources for education remained unresolved. The Whitlam government used its powers under Section 96 of the Commonwealth Constitution to attempt to achieve its socialist, egalitarian objectives by transforming education. Commonwealth support for, intervention in, and influence on education increased dramatically. The reaction of the states was mixed; the added support for schools was welcomed, but inter-government relations became strained with the growing centralisation of power. The Liberal policy statement of 1975 (43) expressed its opposition to "central authoritarian control" and "heavy handed interference". The party reasserted its commitment to "co-operative federalism" in which the powers of Commonwealth, State and Local Government are distributed. But the policy did not preclude a liberal government from evoking special purpose grants "to initiate programs in agreed areas of national need".

In theory, the policy of "co-operative federalism" promised a shift in power from Canberra and an improvement in inter-government relations. In reality, the strong conservative reaction to the Whitlam era and the serious economic conditions of the time militated against any loosening of central authority.

Given the structure of education in Australia and the stated policy of federalism, Senator Carrick, as Minister of Education, needed to secure the support of the state ministers for Commonwealth initiatives. These initiatives included the establishment of the AEC Working Party on Teacher Supply and Demand; National Inquiries into Teacher Education and Education, Training and Work; the School-to-Work Transition program; and ASSP. These programs reflect the Fraser government's ideological commitment to small government, to reducing the public sector, and to re-establishing the demands of the economic system as paramount in defining the purposes of education.

Fears of central intervention and ideological differences notwithstanding, Senator Carrick skilfully steered through the AEC the programs central to his government's purposes, linking these repeatedly with the need to improve standards. Whatever their own position, the State Ministers and Directors-General interviewed as part of this evaluation were generally of the view that he sincerely believed that national monitoring of standards was needed. But why? The picture which emerges from the limited evidence available (44) is not clear, but the notion was politically, managerially and educationally attractive.

Politically, a good deal of heat had been generated during the standards debate. There were advantages in acting in what seemed to be a rational and responsible manner. The introduction of a national monitoring program like NAEP promised to settle the disputes and to satisfy the demand for an objective indicator of the educational health of the nation. The introduction of an accountability mechanism sponsored by the AEC also had the advantage of reducing the dependence of Ministers on their Departments for information, thereby reaffirming conventional notions of Ministerial responsibility. There may have been, however, a hidden agenda. The program could serve to strengthen

the hand of the central government in pressing its educational priorities.

In recent years, the growth in size and complexity of bureaucracies has led to a growing conviction of the need for more rational and business-like management systems. Senator Carrick seems to have been convinced that the technology existed or could be developed which would facilitate co-operative planning, policy formulation and evaluation. By defining and focussing attention on the basic objectives of state education (45) and constantly monitoring performance, national assessment promised to increase the cost-effectiveness and responsiveness of the education system.

Educationally, the hope was that standards would be improved. This hope was the product of the long standing belief in testing as a quality-control mechanism and the shift in the ideology of public administration referred to above. Senator Carrick expected that areas of national weakness in the basic skills and sub-populations in need of additional assistance would be identified. The presumed connections between national testing on the one hand and improved student performance on the other were not, however, made explicit.

Whereas Senator Carrick's advocacy of national assessment began in 1976, it was not until the February 1979 AEC meeting that the Webster Committee Report was tabled and its recommendations discussed. It was at this meeting that Senator Carrick, supported by Mr. Thompson (Victoria), began to press for the authorisation of a national testing program. The AEC is reported (46) to have asked the ACER to proceed with development of tests to help teachers and schools to diagnose students learning difficulties and monitor progress. The AEC established a Sub-Committee comprising Mr. Swan (Director-General of NSW as Chairman), Dr. Keeves (ACER) and Mr. Dunn (ERDC) charging them with the task of preparing a more detailed plan. At the AEC Meeting in June, 1979 approval in principle was given for the development of a School Achievement Testing program. The Sub-Committee was asked to develop a "comprehensive set of proposals to enable the Council to decide on a program of research and development in this regard at its next meeting in Perth" (47).

In October 1979, immediately prior to the AEC meeting in Perth, the AEC Standing Committee considered the final version of the ACER proposal. This committee acts in an advisory capacity to the Ministers and includes the heads of State and Commonwealth Education Departments, tertiary and technical education authorities. The Directors-General on the Committee had anticipated a proposal with a structure for the development of Progress and Review tests which included a concept of sharing between the States and the ACER. They believed that there was a need for diagnostic tests of the Progress and Review type and that these would be of considerable value to schools and teachers. They were not opposed to state monitoring of student performance (several states had basic skills testing programs) but feared a national testing program would have an undesirable backwash effect on the curriculum, and that the results would be used once again in a political attack on education systems. The Directors-General then were unanimous in their support for the

Progress and Review proposals and opposed to national Survey testing (48). The Chairman of the Standing Committee (48) reported that some other members of the Committee (notably heads of Commonwealth Departments) would have qualified this position but, given that it was a schools matter, it was the Directors-General recommendation which was important. Almost always, the Ministers accept the recommendations of the Standing Committee. The Standing Committee endorsed the first part of the recommendations of the Sub Committee in relation to proceeding with "Progress and Review" testing. The following day the ACER proposal and the Standing Committee recommendations went to the AEC. For the reasons advanced earlier, the Commonwealth and Victorian Ministers argued in favour of a national survey. On the other hand, some State Ministers, acting on advice from their Directors-General, are reported to have spoken against the proposal. Still others, while aware of their Director's opposition to survey testing, came out in favour of national monitoring of literacy and numeracy. The reasons for support or opposition are undoubtedly complex. In some cases, Ministers may have been convinced that state monitoring procedures already in operation were more appropriate than the proposed national program. In other cases, support for the proposal may have been seen to be consistent with government policy. For example, SA Liberal Education Policy had pledged in August 1979 that all schools would take a common core of instruction in reading, writing and arithmetic "so that there can be a common evaluation of student performance" (49).

In a close vote, the AEC decided to commission ACER to compile a series of tests for national sample surveys over the next five years. The Council also endorsed proposals for the co-operative development of Progress and Review tests. It was agreed that the survey testing would be done by ACER in co-operation with the State Departments and those reponsible for non-government systems. Budget ceilings were decided for Survey testing, the estimates being $90,000 for the current (1979-80) financial year and $140,000 per annum for the successive four financial years. Half the funds were to be provided by the Commonwealth and the rest from the states on a per capita basis. It was anticipated that survey testing of literacy and numeracy would begin in October, 1980, and that about 1000 students aged 10 and 14 would be tested in each State on an annual basis for the next five years. The Chairman (50) reported that only national results would be made public; the results for individual schools would be provided only to those schools, and State results only to the Ministers and Departments of Education in those States.

THE ROLE OF THE STEERING COMMITTEE

One issue of major importance in commissioned research, development and evaluation projects has been that of supervision and control. Frequently, the organization commissioning the work seeks to call the tune, while the reseachers strive to retain their independence. As an independent, national research organization, ACER has sought to maintain its autonomy in conducting research. At the same time, it has attempted to secure long-term contracts and commitments from governments in accordance with its developmental plans. The ACER proposal suggested a management structure which

included a small Steering Committee (two DGs, two ACER staff and the Chairman of ERDC) responsible to the Ministers for the overall direction of the project, a large Reference Committee, and small Technical Committees for each curriculum area assessed.

From the perspective of the Directors-General, a high degree of AEC-systems control over the project was essential if backwash effects on the curriculum were to be avoided. The Directors-General were not happy then with the management structure proposed by ACER and were successful in their efforts to change the membership of the Steering Committee responsible for directing the project. The Committee which was established included representatives of all eight government systems and the Chairman of ERDC, and was chaired by Mr. D.H. Swan (Director-General NSW) with Dr. Keeves as an ex officeo member. The tensions between the key systems and ACER representatives on the Steering Committee (which had been evident between Keeves and Swan over the ACER proposal) immediately surfaced. At the first Steering Committee meeting in November, 1979, the State representatives differentiated between the Progress and Review testing program, which they supported, and the Survey program, which they wished to keep separate and as low key as possible. They were particularly concerned about the control over, and the uses of, the national survey. It was feared that if national survey and Progress and Review tests were drawn from the same set of items, the tests in effect would determine the curricula of schools. Dr. Keeves, on the other hand, argued that from ACER's point of view, Progress and Review and Survey tests should be developed as part of the one co-ordinated program. ACER wanted the development of a national monitoring program linked with the 1975 study. It also sought to expand the scope of the program to include speaking skills, listening skills, study skills and science. ACER argued that there would need to be a commitment to the survey testing program for a minimum of five years. In the debate which followed, it was recognized that a statement indicating the key items of agreement between the AEC and the ACER was needed. A Technical Sub-Committee was established to look at such matters as sample size, types of test items, test objectives and methods of reporting results.

In December 1979, the Technical Sub-Committee met to identify issues requiring resolution by the Steering Committee on behalf of the AEC (51). The issues identified were numerous, reflecting the systems' uncertainty about the intentions and expectations of the Ministers and the conflicts among the various interests involved in the initiation, development and implementation of the AEC project. The most substantial of these issues related to the underlying purposes of the study; the breakdowns and comparisons to be made in the survey results; the scope and frequency of testing; access to the data; reporting of results; and the controls needed to minimize the effects on the curriculum, as well as misuses of the results (52). Exactly what was intended by the AEC was not spelled out in detail. The statements made by the AEC and expectations of the Ministers were open, therefore, to different interpretations. Dr. Keeves took the view that, as there was no formal contract forthcoming, the AEC Press Release and decisions reported in the Minutes of the October 1979 AEC meeting should be regarded as a contract for work. In his account, the AEC had endorsed the proposals advanced by ACER and had agreed to

a budget ceiling of $90,000 for the first year and $140,000 per annum for the successive four years. The Ministerial statement and decisions, however, also speak of co-operation between ACER, the State Departments and non-government systems, with the overall direction of the project in the hands of the Steering Committee. The Directors-General took this to mean that the Steering Committee, not the ACER, was to assume the major reponsibility for directing the testing project.

The Steering Group and its Technical Sub-Committee addressed themselves to the questions of sampling, frequency of testing and the use to which background variables would be put, finding it essential to clarify the purposes of the survey before agreeing upon answers. A draft document on purposes, prepared by Keeves and Donovan (ACT Schools Authority), was examined by the Steering Committee and amended, without receiving final approval. The committee did decide, however, that a general description of the national survey program should be prepared for the information of the various interest groups who could legitimately be expected to have concern for the program.

From the question of sampling, other questions followed; for example, what data would Ministers receive? Would the data be national only, or refer to their own State, aggregated data only, or data on individual and identifiable schools? Other questions arose: would Directors of systems receive data pertaining to their system? How valuable is light sampling? As it is a national survey, why sample within States? Is light sampling an economic necessity?

While agreement was reached on light sampling, there was conflict and confusion over analysis of sub-populations. The ACER wished to replicate the analyses undertaken in 1975. The Schools Commission had hoped that results of the survey would assist them to analyse effects of some of their programs (eg. disadvantaged schools, migrant education), and to assist their decisions on allocation of resources. Overall, however, the fear of misuse of results prevailed over the possibility of reaping benefits through comparisons between sub-populations. Past attacks on schools and education systems by politicians and employer groups given prominence in the daily press, had so affected systems morale that comparisons between States and/or systems were to be avoided at all costs.

Concern over the 1975 sampling of Aboriginals and the potential reactions of disadvantaged groups, contributed to the pressure to remove the "ethnic" category. Politically, breakdowns by ethnicity were unacceptable to the systems representatives. The ACER took the position that such breakdowns were feasible and useful, and that comparisons should be made between student performances in 1975 and 1980 in several broad language categories. The systems representatives believed that reporting on ethnic groups was undesirable and that in no case would there be sufficient numbers for a meaningful analysis to be possible. In 1975, the numbers of children in all but the English only category were such as to make interpretation of trends in performance difficult. Ultimately the Steering Committee decided that no breakdowns by language group would be reported. Other potential problems in sampling and interpretation

of data were also raised. For example, South Australia pointed to differences between the States in year levels of children of the same age and to variations in the implementation of policies of integration of 'special education' groups in normal schools. Both of these factors would affect the nature of populations sampled within and between States.

The fear of central (ie. Commonwealth) control of curriculum was also voiced, and arguments that light sampling, the nature of the tests and restrictions on the information would minimize backwash effects only went a small distance towards dissipating this fear. Some committee members were also unsure if tests originally designed to measure learning difficulties were valid as measures of standards and feared that annual testing using the kind of items contained in the tests could come to define Australian primary curricula. The form, audience and responsibility for reporting of the results proved to be a source of concern. The ACER representatives declared their unwillingness to be placed in a position wherein their reports might not be released or released only in a substantially ammended form. While it was generally accepted that the major report and a summary report would be prepared by ACER and released through the AEC, the systems representatives were concerned that the ACER Reports would not give an adequate overview of the position adopted by the opponents of ASSP or of the debates within the Steering Committee. They were strongly of the view that press releases should be the responsibility of the AEC through its Standing Committee. In fact, the summary and final reports prepared by ACER were released by the AEC; and the press release was written on behalf of AEC by an officer of the Commonwealth Department of Education.

On another front, tensions also developed over the Progress and Review testing, fundamental difficulties arising out of the insistence of some states on having the products attuned to their requirements, and out of the suspicion that due consideration might not be given to the work done already by some states. Two management models were proposed: one viewed ACER as the co-ordinating body, the other proposed a representative management group which would leave greater control in the hands of the States. The latter model prevailed.

The Steering Committee consulted with representatives of non-government schools (independent and catholic systems), teachers' unions and parents' organizations. The fact of a National Survey was grudgingly accepted, although doubts about the usefulness of the program and concerns about misuses of the data were expressed, particularly by the Australian Teachers' Federation.

At both national and state meetings of interest groups organized by members of the Steering Committee, consensus was reached on a number of issues. Like their counterparts from the government sector, the representatives of non-government systems were firmly opposed to any comparisons being made between systems at a national or state level, and to any possibility that individual schools might be able to make their results public. They were also agreed that the number of associated background variables (especially ethnicity and socioeconomic status) should be kept to a minimum. Most interest

groups lamented the fact that they were not represented on the Steering Committee.

Before the June 1980 meeting of the AEC, the Steering Committee, having consulted other groups, felt obliged to re-think their decisions, particularly in regard to schools' and individuals' rights to decline to participate in the survey, and the availability of test results to various bodies from Ministers and Directors through to principals and teachers. It was agreed that the project be known as Australian Studies in Student Performance to indicate continuity with the 1975 ASSP (Australian Studies in School Performance) testing program, but to indicate a shift from school to student in the statement of the purpose of the survey.

Between October 1979, when the AEC decided to proceed with a national survey, and October 1980, when the ASSP tests were administered, the scope of testing had been limited to minimum competency of 10 and 14-year olds in the performance of specific tasks in areas of reading, writing and numeration. The extension of the program to other curriculum areas (as sought by Victoria, for example) and attitudes to reading (suggested by ACER) was disallowed. Light sampling only was agreed to, and Victoria and N.T. which wanted total cover eventually retreated from this position.

In this period it became clear that the interests of the ACER, as the national research organization responsible for carrying out the project, and the State Departments were in conflict. The systems were cynical about the purposes of the project and the motives of its proponents. Hence they sought to prevent the possibility of invidious comparisons being made between states, government and non-government systems and ethnic groups, a concern shared by the representatives of non-government schools and many interest groups. The Samuel article in the Bulletin had all too clearly revealed how readily invalid causal inferences are made on the basis of comparative descriptive data. The ACER questioned the propriety of the systems' efforts to intervene in a project approved by the AEC. From the ACER's perspective, the management situation was "virtually unworkable" (53). On the other hand, the AEC and its advisors were concerned to ensure adequate measures were taken to minimize the potential dangers of national testing.

The 41st meeting of the AEC was held in June 1980 in Sydney with Mr. Landa, the NSW Minister, as Chairman. The recommendations of the Steering Committee along with correspondence from the ATF expressing its concerns were tabled and discussed at the Standing Committee and the AEC meeting. Fourteen recommendations were endorsed by the AEC, with some modifications being made to recommendations relating to the breakdown of results. The Ministers were to receive a data tape with sufficient information to allow each state or territory to identify its own data, but the Ministers agreed not to use the data to make comparisons between systems. Potential background data were to be collected on students to ensure comparability with the 1975 samples, and analysis of the data at the national level was to take into account, if possible, sex, ethnic and urban-rural differences. It was agreed that the development of the Progress and Review tests was to be

a separate and independent task from the ASSP survey tests. Approval was also given for the present evaluation study.

At the June 1980 meeting, the main interest of the Directors-General was to minimize the dangers of national assessment, thinking this was the best that they could hope to accomplish. Changes in the membership of the AEC (eg. Mr. W. Fife was now Commonwealth Minister) and the emergence and co-ordination of pressure groups opposed to the national survey changed the nature of the debate in the Council. The Ministers decided that the 1980 ASSP testing program should proceed as already defined, but that the nature and frequency of testing for subsequent years (originally planned for 1981, 1982 and 1983) be re examined by the Steering Committee and referred to the AEC for consideration. As the official statement restricted the purposes of the program to an information-gathering exercise, the question of access to and dissemination of data remains an important issue.

While AEC policy regarding the future of ASSP is uncertain, the Senate Standing Committee on Education and the Arts, after inquiring into the effectiveness of Australian schools in preparing young people for the workforce, has come out in favour of continuous national monitoring of standards of literacy and numeracy on the grounds that:

> criterion-referenced assessments, administered impartially, will be of great assistance to the teaching profession in the longer term as deficiencies in the education are detected and remedied and the results fed back to teachers (54).

The Senate Committee, after listening to evidence from employers, tertiary institutions and school authorities, concluded that unacceptably large numbers (estimated at between 20 and 25 per cent) of students were leaving school lacking in the skills of literacy and numeracy needed to compete effectively in the job market. In the Committee's view, the debate on declining standards is "irrelevant" to the major issues and an "unfortunate waste of resources and expertise". Because of changed societal expectations over time which have in turn influenced educational priorities and the size and nature of the student population, the Committee believes that we will not be able to determine whether in fact standards have shifted over time. The report focuses instead on the degree to which there is a mismatch between standards being achieved in today's schools and societal requirements, but skirts over difficulties in defining standards and societal requirements.

OPPOSITION TO ASSP

Prior to the decision of the AEC in October 1979, there was little public debate on the issue of national assessment, the opposition being unprepared and unco-ordinated. The ATF expressed its concern that the unions had not been consulted by the AEC. At the January 1980 ATF Conference a decision calling for consultation and discussion on the national testing project was circulated to all Ministers. Soon after, the ATF met with Mr. Fife (the new

Commonwealth Minister) to express its concerns and reservations.

The groups opposed to national testing began to co-ordinate an attack on the AEC project. The coalition included the teachers' unions, state school parents' organisations, the Curriculum Development Centre (particularly Dr. Skilbeck) and some educationists.

Much of the substantive criticism stemmed from Dr. Skilbeck who, as Director of the Curriculum Development Centre, was critical of the rationale underlying national testing and very concerned about the impact of the program on schools. In his view, ASSP was likely to narrow teaching uduly and hinder the development of curricula which engage the interests and energies of young people. Dr. Skilbeck argued that educational standards ought not to be seen as fixed points on a quantitative scale derived from an "objective" test; rather, they are complex, qualitative and value-saturated definitions of what educators are striving for. Interpreting evidence relating to changes in performance measures is difficult, and doubly so in the absence of data about changes in student populations, school programs and society. Attempts to monitor standards then run grave risks if they are not accompanied by a "resolute and effective campaign of public education in the meaning of national data". While Dr. Skilbeck is concerned that the available evidence does not warrant assertion that standards have declined, he accepts that there are legitimate areas of public concern about schooling. Accordingly, Dr. Skilbeck has outlined the measures that he believes need to be taken to improve the quality of schooling and public debate about education (55). The CDC believes that there are better ways of maintaining educational standards and having schools account to society than National testing. The alternative proposed by CDC is for well-developed, individual school monitoring programs with agreed reporting procedures acceptable to the community at large. Skilbeck's concerns about implications for the curriculum were shared by Lawton (56), and Fensham (57). Like Skilbeck, Ransley and Little (58) questioned the validity of the results and their information value. The unions questioned the cost of the project and expressed concern over the political misuse of the results. The grounds for objection were outlined largely in union journals and the CDC Core Curriculum document (59). The press coverage given (60) to opponents of ASSP, however, has been minimal compared with that given to the 1975 survey results.

The major grounds for objection were as follows:

(i) Faulty underlying assumptions. The assumption that national testing would produce "hard, scientific" data on standards and that these would be useful in informing policy and improving practice was challenged. The evidence obtained in so-called "scientific" research in education has often proved shaky and more likely to generate speculation and controversy than certainty (61).

(ii) Doubtful validity. The validity of the tests as measures of "basic" skills is open to question. What is "basic" is a matter of judgment, as are pass-fail "standards".

(iii) Limited value. National testing will tell us mainly what we already know; for example, that children from non English speaking and disadvantaged families have difficulty with reading, writing and numeration tasks.

Even if trends can be established, no explanation or means of ruling out competing hypotheses as to the reasons is possible. National assessment, in effect, confuses symptoms with remedies. Teachers and policy makers will still be left with no clues as to what can and should be done. More significant and enduring results are likely to come from curricula and materials (including Progress and Review tests) based on an understanding of the teaching-learning process and designed to improve the capacity of teachers to make informed and sensitive judgments.

(iv) Harmful side effects. The dangers of negative backwash effects on the school curriculum, political misuse of the results, and social divisiveness are especially great in the current situation, with schools and systems competing among themselves for their share of a declining school population and education budget. Teachers will be encouraged to prepare students for batteries of tests instead of focussing their attention on the needs of the child and the requirements of society. Developments in the UK and USA suggest that, despite reassurances and planned safeguards, minimal competency and national assessment programs are likely to mushroom into large scale operations which have the effect of turning schools into "skills factories" (62). The attacks on schools following the 1975 testing were ill-informed and largely ideological. These resulted in a serious loss in morale amongst teachers and in confidence in the school system within the community. The 1980 program gives every opportunity for a renewed attack and little hope of increasing public understanding of matters relating to schooling and student achievement.

(v) Cost. The survey testing program as proposed by ACER (costing $650,000 over a five year period together with the costs of time of all involved) is a waste of scarce financial, and human resources.

(vi) Distancing. The national survey provides a mechanism whereby Ministers can distance themselves from their responsibilities; systems and schools become the scapegoats, while any serious consideration and open discussion of the weaknesses of schools, the problems of social disadvantage and the youth unemployment crisis is averted (63).

In mid 1980, the ATF wrote to the Ministers asking them to consult and to call off the testing scheduled for October. The Victorian Minister indicated that the request was too late and that adequate consultation had already taken place. The correspondence between ATF and Mr. Swan was, however, tabled at the June, 1980 AEC meeting. At the June ATF executive meeting, it was moved that ATF should announce a boycott, but the motion was not carried. At the September meeting, ATF received a report from affiliated unions (South Australia and Tasmania are not members of ATF) favouring a boycott (64). The ATF decision, however, was too late to be fully effective.

While the boycott received wide publicity in some states (Victoria, Queensland, Tasmania) and the Territories (particularly N.T. where blanket testing was being pushed), in others it was either very late (in NSW a circular was sent to schools in the week of the testing) or non-existent (SA).

The testing went ahead as scheduled in October, 1980.

CONCLUSION

In this chapter we have sought to explain why ASSP was introduced and to chart its development up to the point of testing, locating and placing in a broader social, economic and political context the events, individuals and groups involved in the decision-making process. Our analysis reveals that the publication of the 1975 survey conducted for the purpose of identifying specific learning disabilities co-incided with the intensification of concerns over the quality and cost-effectiveness of education. Senator Carrick made evident his determination to acquire more systematic feedback about standards from the school sector and the ACER indicated its willingness to develop the appropriate technology. In its early phases, the proposal was not, however, the subject of open debate - with some unfortunate consequences.

The decision to proceed with the testing program was made against the advice of the Directors-General and without consulting the unions or securing their support. Politically, ASSP became a powder keg. The subsequent attempts by the Education Departments to minimize its potential dangers and the Unions to destroy it seem destined to limit whatever impact, good or bad, the project might have had. The suspicions and tensions generated seem to have inhibited frank and open discussion of such issues as the quality of education, the accountability of schools and the contribution of assessment to policy and practice. In the chapters which follow, the consequences of decisions made and actions taken in the development of ASSP are examined in greater detail.

REFERENCES

1. Schon, D.A. Beyond the stable state: public and private learning in a changing society. London: Maurice Temple Smith, 1971.

2. Cadman, A.G. (Chairman) Learning difficulties in children and adults. Report of the House of Representatives Select Committee on Specific Learning Difficulties. Canberra: AGPS, 1976.

3. Keeves, J.P. and Bourke, S.F. Australian Studies in School Performance: Vol.1, Literacy and Numeracy in Australian Schools: A first report. ERDC Report No.8. Canberra: AGPS, 1976.

Bourke, S.F. and Lewis, R. Australian Studies in School Performance: Vol.II, Literacy and Numeracy in Australian Schools: Item Report. Canberra: AGPS, 1976.

Bourke, S.F. and Keeves, J.P. (eds.) The Mastery of Literacy and Numeracy: Final Report. Australian Studies in Student Performance: Vol.III, Canberra: AGPS, 1977.

4. Cadman op. cit.

5. College Entrance Examinations Board. Report of the Advisory Panel of the Scholastic Aptitude Test Score Decline. New York: College Brand, 1977.

6. Hopkins, A. The School Debate London: Penguin, 1978.

7. Cadman op. cit. p.14

8. Mr. I.C. Spicer (Secretary Victorian Employers Federation) in The Australian, 1/3/1975.

9. Australian Council for Educational Standards, Statement of Principles. ACES Review, 1973, 1(1).

Just, F.P. The case for examinations. Parkville, Victoria: ACES, 1973.

Barcan, A. (ed). The quality of the curriculum (Articles from ACES Review). ACES, 1982.

10. Peter Samuel. Australia's education scandal: we're turning out millions of dunces. The Bulletin, May 15, 1976.

11. Dr. K. McKinnon strongly defended teachers in his submission to the Senate Committee on Education and the Arts.

12. Keeves, Matthews & Bourke op. cit. p.45.

13. cf. Fraser's Philosophy. Edited text of a speech to the S.A. State Council of the Liberal Party. The Advertiser, Dec. 17, 1980, p.5. School fundings issues primed to explode. The National Times, Dec.7 to 13, 1980, p.5.

14. Australia, Senate, Debates, 1976, Vol.567, p.1176.

15. Australia, Senate, Debates, 1976, Vol.567, pp.1176-7.

16. Australia, Senate, Debates, 1976 Vol.568, p.1784.

17. Australia, Senate, Debates, 1976, Vol.569, p.1294.

18. Australia, Senate, Debates, 1976, Vol.567, p.2009.

19. Australia, House of Representatives, Debates, 1976, Vol.101, p.1846.

20. Australia, House of Representatives, Debates, 1976, Vol.102, pp.3648-51.

21. Australia, House of Representatives, Debates, Vol.102, p.2999.

22. Australia, Senate, Debates, 1977, Vol.571, p.12.

23. Australia, Senate, Debates, 1977, Vol.579,pp.1703,1757.

24. Australia, Senate, Debates, 1977, Vol.578, p.600.

25. The ACER made a detailed submission published later as: Keeves, J.P., Matthews, J.K. Bourke, S.F. Educating for Literacy and Numeracy in Australian Schools. Melbourne: ACER, 1978.

26. Australia, Commonwealth of. Education, Training and Employment. Report of the Committee of Inquiry into Education and Training. Canberra, AGPS 1979, p.124-26.

27. Editorial, The Australian, 23/10/1980.

28. The Australian, 15/1/1980.

29. Adams, R. Accountability in Australian Education. In Australian College of Education. New Directions in Australian Education. Carlton, Vic.: ACE, 1976.

30. White, M. Post school education and employment in Western Australia during the 1930's and 1970's. Paper presented at AARE Conference, Sydney, 1980.

31. Bessant, B. and Spaul, A. Politics of schooling. Carlton, Vic. Pitman, 1976.

32. Communication from Senator Carrick, 17/9/1980.

33. ERDC. National Assessment of Educational Progress. ERDC Report No.26 AGPS, 1980.

34. op. cit. preface.

35. op. cit. p.37.

36. Notably Ms. J. Blackburn and Dr. M. Skilbeck.

37. Jackobson, J. A summary of the research into reading standards of Queensland Grade five students, 1933-1977. Brisbane: Queensland Institute for Educational Research, 1978.

38. Start, K.B. and Wells, B.K. The trend of reading standards. London: NFER, 1972.

39. K.B. Start. The test of testing. Address to the South Australian Institute of Educational Research, April, 1981; Interview, April, 1981.

40. Interview with Hon. A. Hunt and Hon. N. Lacy, 6-10-80.

41. S.F. Bourke. The development of school achievement tests. Hawthorn, Vic., ACER August, 1979.

42. J.R. Kerr. Bureacracy and society. In Parliament Bureacracy, Citizens: Who Runs Australia? Sydney: Angus & Robertson, Australian Institute of Polical Science, 1972.

43. The Liberal Party Federal Secretariat, Federalism Policy, September, 1975.

44. The evaluation team did not have access to the confidential records of AEC meetings and had to rely on information given in interviews and public documents.

45. Cohen and Haney [Minimum competency testing and social policy. In R. Jaeger & C. Hehr (eds.) Minimum competency achievement testing. Berkeley: McCutchan, 1980] point out there has been a long tradition in government policy of assuming responsibility for minimum levels of health, education and social welfare.

46. Introductory remarks of the Chairman, AEC Steering Committee meeting, 8-2-80.

47. Press statement by Chairman of the AEC, Hon. P. Jones, W.A. Minister for Education, 30 June, 1979.

48. Dr. D. Mossenson, DGE Western Australia, Interview, November, 1980.

49. Liberal Party Education Policy, H. Allison, Shadow Minister of Education, August, 1979. Also in August, 1979, Dr. Keeves had talked at Mt. Gambier East High School on "Standards and Youth Unemployment", had met Mr. Allison (who is member for Mt. Gambier) and subsequently sent him the ACER report on the 1975 testing.

50. AEC Press release, 26 October, 1979.

51. Interviews with members of Technical Sub Committee - P. Tillett, M. Angus, L. Blazeley.

52. Victoria and Northern Territory Ministers were keen to have systems- wide rather than light sampling.

53. Dr. Keeves, interview, 10-12-1980.

54. Senate Standing Committee on Education and the Arts. Preparation for the Workforce. Canberra: Australian government Printing Service, 1981, p.23.

55. Skilbeck, M. Perspectives on performance and prospectives for future initiatives. Quest, 1978, 24, 33-38.

56. Lawton, D. The politics of the school curriculum. London: Routledge & Kegan Paul, 1980.

57. Fensham, P. National testing: a counter productive contribution in the educational task. Paper Section 22, ANZAAS, May, 1980.

58. Ransley, W. and Little, G. Testing standards in basic skills. Education News, 1979, 16(10), 20-24.

59. CDC. Core Curriculum for Australian Schools. Canberra: CDC, 1980.

60. The Age, 31-3-80; Education News, National Times op cit.

61. This point has been made with respect to other major studies (eg. Coleman) by D. Cohen and M. Garet (Reforming educational policy with applied social research, Harvard Education Review, 1975, 45(1), 17-43).

62. cf. H. Broudy. Impact of minimum competency testing on the curriculum. In Jaeger and Kehr op cit; P. Fensham op cit.

63. Ginsburg, M. et al. Teachers, The Great Debate and Education Cuts. Westminister Studies in Education, 1979, 2. 5-33.

64. The Sydney Morning Herald, Sept. 16; The Australian, Sept. 22, 1980.

CHAPTER THREE

EXPECTATIONS AND OPINIONS OF ASSP

Mal Hewitson and Neil Baumgart

INTRODUCTION

In Chapter Two, discussion focussed on the historical origins of the ASSP project and the rationale underlying its evolution up to the October, 1980, testing period. Mention was made, too, of the uncertainties experienced by various groups within the Australian community regarding the purposes of the project and whether ASSP was a cost-effective investment of funds.

The purposes of this chapter are twofold. In the first place it asks what the community can expect of ASSP given the restriction of testing to aspects of literacy and numeracy using a light national sample of 10 and 14 year olds. Secondly, the chapter seeks to identify concerns about, and expectations of the ASSP project at the time of testing.

For the former purpose, the literature on national assessment generally was examined. Particular attention was paid to testing in areas of literacy and numeracy at the national level, and to discussion on what a testing program with limitations similar to ASSP might or might not achieve.

To obtain information on concerns about and expectations of ASSP (the second purpose), several strategies were used. In July 1980, the research team met formally with representatives from a number of national organisations having a direct interest in the ASSP project and its outcomes. The organisations included the Australian Education Council (AEC), the Education Research and Development Committee (ERDC), the Schools Commission, the Curriculum Development Centre (CDC), the Australian Teachers Federation (ATF), the Australian Council of State School Organisations (ACSSS), the National Council of Independent Schools (NCIS), and the National Catholic Education Commission. Although the main purpose of this meeting was to guide subsequent evaluation strategies, many of the issues raised during the meeting are included in the present chapter.

The evaluation team also developed a survey questionnaire which canvassed the level of knowledge about the ASSP project and views on and expectations of it expressed by a reasonably broad cross-section of professional groups of educationists and other interested bodies. This aspect of the evaluation study was prompted by the uncertainties surrounding the nature of the ASSP project. It was argued that if the primary aim of the ASSP project was "to gather information on achievement in aspects of literacy and numeracy...and to make this

information available to the public" (1) then the level of public awareness of the project and its purposes at the time of the testing period (October, 1980) should be reported, along with public expectations of what the project would or could achieve. Moreover, it was thought desirable to evaluate and explain public reaction to publication of the findings in light of the public's prior knowledge of the project (or lack of it) and expectations.

For these reasons, then, a number of groups were surveyed in October 1980, just after the administration of the ASSP tests. They included education system administrators, teachers' federations/unions, professional associations of principals and teachers, parent organisations, employer groups, and other organisations expressing views about standards of literacy and numeracy and/or national assessment of student progress. Their respective concerns about, and expectations of, the ASSP project are summarised in the latter part of the present chapter.

WHAT CAN BE EXPECTED OF ASSP?

In Australia, the United States, the United Kingdom and elsewhere, community concerns about standards, needs for information and demands for accountability led to the introduction of national assessment programs, the stated purposes of which include providing information for the public. Such programs, and public announcements about them, generate expectations about what information will be provided, what the results will show and how these will be used, and at the same time generate concerns about the impact of the tests on the curriculum and possible misuses of the data.

Thus one set of concerns leads to the belief that "hard evidence" of changes in standards will be provided and concerns about declines will be confirmed or refuted, and to the expectation that testing will lead to policies and changes in schools which will result in improvements in literacy and numeracy and in the quality of education. On the other hand, other concerns focus on the backwash effects of testing on the curriculum and teaching and the expectation that the results may be used to legitimise budget cuts and controls on schools.

The purpose of this section is not to discuss arguments for and against projects such as ASSP per se. Rather it is to discuss those factors which complicate the answer to an apparently very simple question, namely, what is a national assessment project such as ASSP designed to achieve? Complications arise because decisions need to be made on the content areas to be tested; the range of student performance to be tested; the standards of attainment to be used to determine success and failure; the selection of students to be tested; the additional data to be collected; the interpretation of findings; the form in which the results are to be made available; and to whom they are to be released.

These issues and the implications of decisions taken on them are addressed below.

IMPLICATIONS OF DECISIONS ON CONTENT AREAS TO BE TESTED

The decision to include only specified areas of literacy and numeracy in the ASSP testing program immediately limits what the project can say about the minimum competencies of students in Australian schools. Clearly, levels of student performance can be established only in the content areas tested. Nothing is established about student performance in the many other untested content areas, or about student attitudes, or about social, moral and ethical standards of student behaviour, or about growth in physical skills; yet all of these may legitimately be associated with desirable aims and outcomes of schooling.

IMPLICATIONS OF DECISIONS ON THE RANGE OF STUDENT PERFORMANCE TO TESTED

In any given content area, student achievement normally ranges widely. Until recently, most achievement tests given to students in Australian schools have sought to discriminate among students who perform poorly, students who perform satisfactorily or well, and students who perform very well or excellently on the test. That is, they test for a relatively wide range of students' achievement levels, even though they rarely test for the full range of achievement.

The ASSP project did not utilise achievement tests of this kind. It did not seek to test a wide range of student performance in reading, writing and number work. What ASSP did was to attempt to separate out only those students whose performance on the tests is so poor that they must be classified as failing to attain those basic skills of literacy and numeracy which are "required for effective functioning in their group and community" (2). To attain this aim, the ASSP project utilised criterion-referenced tests of minimum competency based on the lowest standards of knowledge and skills expected of all target students at the time of testing.

The decision to focus on minimum competencies means that the tests tell nothing about changes in performance at other levels. The point is sufficiently important to warrant an analogy. A society concerned about its poor might monitor the percentage of the population below the poverty line. Such a survey gives no indication of levels of wealth in the community. Similarly, to monitor the performance of the weak in basic skills gives no indication of the standards of the average and above average: changes at these levels remain undetected.

IMPLICATIONS OF THE STANDARDS OF ATTAINMENT USED TO DETERMINE SUCCESS AND FAILURE

There are two main methods of determining standards of student attainment in tests and examinations. The first method involves comparing the student's score with a standard of attainment which is pre-determined. This method (criterion-referenced testing) compares student performance with a pre-specified "cut-off" (criterion) or standard of attainment. The other main method of determining standards of student attainment is to compare all student scores with

each other by using the average of all the scores as a bench-mark. This method (known as norm-referenced testing) assumes that average student ability across a large population of students (e.g. all year 12 students in a particular state) varies little from one year to the next. Hence similar proportions of the population tested receive A's, B's, C's etc (or 1's, 2's, 3's, etc.) each year.

The ASSP project utilised pre-specified criterion-referenced standards based on a definition of "mastery" of basic skills which equated this with an 80 per cent cut off mark. One practical difficulty typically associated with norm-referenced achievement testing is thereby overcome, namely, the problem of developing national tests based on school curricula having marked differences in content both between and within Australian states and territories. Even quite different curricula are likely to involve similar minimum competencies in literacy and numeracy. However, while minimum competency standards are absolute in the sense of being pre-determined, the specification of these standards is fundamentally a matter of judgment. Such judgment is fairly arbitrary, and could be based on a number of different criteria. Burton (3), for example, discusses standards based on theories, on expert consensus and on practical necessities, and points to the limitations of each approach.

Since any individual's judgment may be at fault, it is logical in a project such as ASSP to reduce individual errors of judgment in setting minimum competency standards by seeking a consensus of expert opinion. The expert opinion might come from three sources - the specialised literature on the topic, theoretical specialists and practising teachers in the field. Nevertheless, the minimum standards finally set are based on responses to discrete tasks, and what the community means by "standards" and what ASSP tests assess may not be the same thing.

It is worth noting that the calculation of average scores from ASSP test results is a meaningless exercise: the only useful information is the proportion of students failing to reach minimum mastery standards. Any concern about ASSP results being used to discover whether students are "up to standard" or "below standard" may be very misleading if by "standard" is meant some standardised norm (e.g. the average score of a large population of students) such as those associated with I.Q. tests or much achievement testing. If, on the other hand, "standard" is to refer to a specified minimum competency standard, phrases such as "above mastery standard" and "below mastery standard" provide clearer meaning.

IMPLICATIONS OF DECISIONS ON WHICH STUDENTS WERE TO BE TESTED

Decisions on which students are to be tested necessarily depend on the information which is required and the uses intended for it. Broadly speaking, tests may be administered to a whole population (e.g. all 10 year olds in normal schools) or to a representative sample of a population (e.g. all 10 year olds born in a particular week or a particular year, or every tenth student on alphabetically listed school rolls). A representative sample is chosen randomly, i.e. chance determines which students in the population are selected

for testing. The size of the sample influences the level of confidence with which findings can be applied, by inference, to the population which the sample represents.

If the information from a national literacy and numeracy testing program were going to be used to make decisions about individual 10 year olds, it would obviously be necessary to test the whole population of students aged 10. If, however, identification of individual student performance is not needed, a representative sample is preferred, since costs then decrease markedly. Under this condition, the size of the sample is determined by how the results are to be used, since comparisons between performances of sub-groups of the population require that each sub-group be adequately represented in the sample.

The present ASSP project adopted a light sampling approach in which approximately one thousand 10 year old and one thousand 14 year old students from each state and five hundred such children from each territory were randomly selected for participation in the project. The sample thus comprises about 3 per cent of each of the populations of 10 and 14 year old students. This means, first, that the results permit reliable statements to be made on an Australia-wide basis about proportions of 10 and 14 year old students in normal schools who are non-masters in the tested aspects of literacy and numeracy; second, that the results may indicate possible trends in student performance between 1975 and 1980; and, third, that state and territory education authorities are in a position to consider implications of the findings for their own state-wide school systems.

IMPLICATIONS OF DECISIONS ON WHAT ADDITIONAL DATA WERE TO BE COLLECTED

The ASSP project permitted male vs female and rural vs urban comparisons. Some information was available on other factors (state, system and language) but it was decided further breakdowns would not be undertaken. Educationally then, the project has limited utility. The results cannot guide teaching strategies nor are they of immediate practical benefit to classroom teachers. At best, the most practical educational outcome of ASSP test results is an indication that diagnostic tests and appropriate remediation of learning difficulties may be required for some (small) proportion of the students in the age group to enable those students to reach minimum mastery standard in the content areas tested. If the ASSP project is repeated every few years or so, provision can be made to make comparisons between students' performances on one testing occasion with student performance on previous occasions, given equivalent tests in the same content areas on all occasions. For example, the 1980 ASSP results can be compared with the results of the 10 and 14 year old students who participated in the 1975 ASSP tests.

Even so, a number of cautions need to be taken into account. Suppose the 1980 results are interpreted as showing an improvement on the 1975 results. It is not possible to identify reasons for the improved student performance as the ASSP project was not designed to search out causes of change. If, however, student performance is tested on a regular basis, then over a long period of time, trends may

emerge which may be interpreted by decision makers and the public as indicative of an improvement (or decline) in the standards of literacy and numeracy of 10 and 14 year old Australian children.

IMPLICATIONS OF DECISIONS ON THE FORM IN WHICH RESULTS WERE TO BE MADE AVAILABLE AND TO WHOM INFORMATION WAS TO BE MADE AVAILABLE

Although the results of national testing programs are of particular interest to political decision makers, there are other audiences who are potential consumers of the new information. These include teacher organisations, parent organisations, statutory bodies such as the Schools Commission and the Curriculum Development Centre, education research and evaluation personnel throughout the country, particular pressure groups such as the Australian Council for Educational Standards, employer groups, and so on. From the condition of confidentiality attaching to the release of results to individual schools and state and territory Ministers of Education, it was clear that the public was to be informed of the findings only at the Australia-wide level of analysis.

There are both disadvantages and advantages in this approach. One of the disadvantages is that local or state pressure groups have no new evidence on which to judge their state's performance in comparison with other states, and hence their possible reactions to ASSP results are de-fused. On the other hand, public reaction in the form of "teacher bashing" or "system bashing" is also considerably weakened. The risk of misuse of reported findings by media distortions may also be greatly reduced as the Australia-wide target tends to be ill-defined, compared, say, to a particular school system. Thus the form in which the results are made available to the public is an important consideration. The challenge for those responsible for the project is to communicate the findings and their interpretation of them to the public using language and terminology that can be readily understood by lay persons. They may, of course, also prepare highly technical reports for the academics and researchers whom they wish to reach. For the general public, however, the dissemination of the main findings in intelligible form may help to forestall the possibility of their misuse.

In summary, there are severe limitations on what can legitimately be expected as outcomes of ASSP, given the very nature of the project. Thus

(i) ASSP does not set out to address questions associated with what should or should not be in the school curriculum.

(ii) The 1980 ASSP project is likely to have negligible influence on the school curriculum and teaching methods.

(iii) ASSP does not seek to establish anything about the kinds of schooling experiences which may affect student performance either favourably or unfavourably.

(iv) ASSP does not attempt to assess the effectiveness of particular inputs into schooling nor directly measure outcomes of

schooling.

(v) ASSP can not directly affect present or future standards in literacy and numeracy (or any other content area).

(vi) ASSP can not establish either what ought to be done regarding standards of literacy and numeracy or how it ought to be done.

(vii) ASSP does not test over a wide range of student achievement.

(viii) ASSP was not designed to diagnose causes for poor performance nor to provide evidence that students are ready to proceed to the next unit of work in the school syllabus.

(ix) ASSP does not distinguish among the majority of students who attain minimum mastery standards nor provide evidence that re-teaching of previous work is desirable or that there are talented students who are ready for enrichment activities.

(x) ASSP does not utilise "average" achievement levels as bench-marks for purposes of comparing individual or group scores with each other, nor does ASSP establish "norms" (average performance levels). It uses criterion-referenced tests and minimum standards and was designed to estimate the proportions of 10 and 14 year olds who reach these standards.

(xi) The confidence which can be placed in the findings of ASSP tests is in large part determined by the size of the probability sample of students selected on a random basis. The sample selected was designed to permit statements to be made on a national basis and on a state or territory basis. In addition, the sex of students and whether they attend urban or rural schools are known. Comparisons can therefore be made between male and female students and between urban and rural students.

Before we leave this section, the matter of comparisons is worth pursuing a little further. Is it possible that some comparisons within the ASSP project are potentially divisive? Can test results be used as a basis for invidious comparisons between individuals or groups of students, between schools, between government and non-government school systems, and between states or territories? Clearly such comparisons are of little use unless they are accompanied by other data which help to explain observed differences.

Sample sizes were chosen to permit comparisons being made using the 1980 ASSP results at state or territory level. Since individual state and territory level results were made available only to the appropriate Minister of Education on a confidential basis, public disclosure of state-level results for comparative purposes was ruled out. It is also worth nothing that participating schools were not identifiable from the results which Ministers of Education receive, and that those schools requesting the results of their own students' performances on the ASSP tests were requested not to make those results public. Under those conditions, even invalid comparisons between schools cannot be made.

Similarly, it might be argued that comparisons between the results of various ethnic sub-groups in the population might lead to social divisiveness. This concern is not applicable, since the relevant information is not included in the major ASSP report.

As noted earlier, lack of comparative data greatly reduces the usefulness of the ASSP project to educational policy makers. For not only are the tests themselves not designed to explain why one class or school or state might differ from another, but also the attention of decision makers is not directed to those classrooms, schools, or states worthy of further investigation in order to discover reasons underlying the different performance of particular student groups.

This is not to say that the results of ASSP which are available to Ministers of Education are of no use at all. Each Minister is in a position to analyse the findings of student performance in his state or territory and compare these with the national picture. In particular, a comparison can be made on each individual item comprising the ASSP tests, and thus action may be initiated in relation to specific aspects of literacy and numeracy. Moreover, school authorities can view students' results in terms of percentages reaching minimum standards in the specified aspects of literacy and numeracy and make global judgments as to whether or not the reported state of affairs is good enough. To this extent, therefore, the primary aim of ASSP "to gather information on minimum levels of achievement in aspects of literacy and numeracy of a sample of 10 and 14 year old students throughout Australia" is met. The purpose - so that "such information can be used to direct attention to areas which may be in need of improvement" - is to a lesser extent fulfilled. In this respect, a strong case can be argued in favour of using progress and review testing programs in all schools to discover areas in need of improvement and students in need of assistance. No national picture, however, would emerge directly from such programs.

ACTUAL CONCERNS AND EXPECTATIONS OF ASSP

In the previous section, a theoretical perspective was adopted in order to examine expectations of ASSP. We now consider actual concerns as expressed by numerous interested parties in the community and report the results of the evaluation team's survey of community expectations of the 1980 project.

Publicly Expressed Concerns

Considerable concern was expressed by various sectors of the community that ASSP testing might well lead to undesirable effects on the school curriculum and teaching methods. Typical of such concern are the following questions put to the evaluation team when it was first formed:

- Will the existence of such studies result in pressures being felt within schools to conform to a fairly rigid national core curriculum to be tested periodically on a national basis?

- Is the hidden motive of the ASSP project finding a "reason"

to institute a vigorous "back to the basics" thrust?

- Won't one immediate outcome be that teachers put more time and effort into coaching their students simply to pass the tests?

It is clear that if national testing were frequent (e.g. carried out every year), if a wide range of content areas were tested, if all or most students in an age group were tested, and if many age groups were tested, there would be good reason for expecting the national testing program to have a substantial impact on school curriculum and teaching methods, somewhat similar to the effects which external public examinations are reported to have. On the other hand, testing a light sample (3 per cent) of 10 and 14 year olds once every four or five years in specified areas of basic literacy and numeracy seems unlikely to have a marked direct influence on the school curriculum and teaching methods; whether testing may indirectly have an impact is less certain. Some argue that the 1975 testing did conribute to a shift in curriculum emphasis towards the "basics" in schools.

Although the future of ASSP is not known in terms of proposed frequency of testing, range of content to be tested, and numbers of students to be tested, to view the current project as the thin edge of a wedge to be used for imposing a national core curriculum on schools is to adopt a position which is not readily defensible.

Nevertheless, the scepticism with which some sections of the community greeted the news of a national testing program is voiced in the following:

> The ASSP will inevitably find that some proportion of Australian children are not literate and numerate. What prior commitment is there from governments to provide appropriate resources to remedy any deficiencies detected through ASSP? (Private communication to the authors, July 1980)

Some sectors of the community also expressed concern about how the results of ASSP would be used. One such concern was that the sample was so small that the ASSP results might lack credibility in the eyes of the public at large. A lack of general acceptance of important findings about national or state-wide trends in student performance in basic literacy and numeracy could certainly negate some of the anticipated values of the project.

The seriousness of this position is revealed in a presidential letter of the Victorian Federation of State School Parent Clubs, in which parents were requested to boycott the ASSP tests as the testing program would...

> Test only a few children in order to measure them or their schools against a so-called standard.
> (The Australian, 26 Aug., 1980, p.3).

Even more importantly, several groups expressed the view that results of the 1975 testing have been reported in a very biased way. They therefore expressed concern that findings from the ASSP project could be grossly misused for political expediency and that there was little opportunity to guard against this possibility.

THE SURVEY FINDINGS

Given concerns such as these, the evaluation team felt it necessary to seek views in a more systematic way from a range of groups at the time the project was initiated. Opinions expressed at this time:

(a) gave information on expectations held for the project;

(b) indicated the extent to which the project was known and understood;

(c) assisted in understanding the support for, and opposition to, national testing; and

(d) permitted a more realistic evaluation of reactions to subsequent publication of results of the testing.

A questionnaire was used to seek opinions from numerous groups and associations including parent groups, teachers' unions, professional associations, employer groups, administrators of independent school systems, selected government agencies, political parties and some groups claiming interest in standards of literacy and numeracy in schools.

The survey cannot claim to be exhaustive of interested groups and, indeed, the list of respondents varied somewhat from State to State. However, a common questionnaire was used and this sought information and opinions on:

- sources of information about ASSP
- perceived arguments for and against ASSP
- expectations of results for the areas of reading, writing and numeration
- likely changes since 1975
- the expected impact of ASSP on policies and practices at a number of levels.

The reactions detailed in this section are based on responses from 60 groups across all States and Territories. A further 13 groups replied but declined to respond to the questions either through ignorance of the project or through a policy decision of the association not to co-operate. Most often responses were completed by an executive officer and were based on previous policy decisions taken by the organisations; others were based on opinions of the executive committee and in a few cases responses were based on a survey of members.

In what follows, opinions expressed are summarised by categories of respondents.

PARENT ORGANISATIONS

Parent organisations contacted include both those associated with government schools and those allied to non-government schools.

The vast majority of those who responded knew of the ASSP project and most had been informed through official channels. The remainder quoted "newspapers" as their source of information.

Opinions about the national testing project ranged from moderate support to vehement opposition. The view held seemed more closely related to State than to any other variable (for example, government/non-government schools). In some States where opposition to ASSP was well organised, parent associations had formed a liaison with other groups such as the association of teachers.

A number of arguments were advanced in favour of national assessment including the need for a quantitative data base to answer criticisms about educational standards; possible uses of findings for future planning or for diagnostic purposes of a global kind; and the need for system accountability. Amongst the arguments most often cited against national assessment were the potential risk of misuse of results for political purposes and the restricted nature of the objectives being tested. Some concern was also expressed about the possibility of teaching for the tests. In addition, the smallness of the sample, the difficulties of measurement, possible intrusions into privacy and the potential danger of having minimum standards seen as optimal standards were also cited as sources of concern.

Most respondents thought the national survey would show areas of strength and weakness and the vast majority predicted little or no change relative to the 1975 results. All respondents indicated they would be seeking a copy of the ASSP report but very few saw any particular benefits for their organisation arising from the project.

Responses to questions on the possible impact of the publication of ASSP findings reflected views held about the worth of the project. Those who felt national testing was worthwhile presented plausible predictions about potential uses by policy makers, employers, teachers and administrators. Those who were sceptical about the project either predicted that it would have little impact or suggested that it would be used by various groups to further their own prejudices.

TEACHERS' UNIONS

The Australian Teachers Federation was opposed to the ASSP project and asked its affiliates to refuse to participate in the testing program. Responses from various unions of teachers refleted this opposition.

In general, the unions advised that they saw no benefit in the present testing project and several gave examples of desired alternative uses for the funds. They indicated that they had been involved in communications with State departmental authorities and, in the case of Victorian unions, with the ACER as well.

The most commonly expressed arguments against the project were as follows:

- the possible misinterpretation of findings, bias in reporting or unwarranted political use of results
- the danger that testing of the "basics" could lead to more rigidity in curricula, more pressure for schools to conform to central control, and less emphasis on broader but no less important aims of schools
- the inability of the project to provide diagnostic assistance to individual learners or schools or to affect the provision of resources to areas of need.

Mention was also made of the confusion between the ASSP project and the use of progress and review tests and attention was drawn to the arguable criteria for item selection.

Most responses from teachers' unions anticipated little or no change in results since 1975 and expressed concerns about the use (or misuse) which could be made of results regardless of what they showed. That is, they feared that any indication of an improvement at the national level could be used as an argument not to provide additional resources for remedial work. On the other hand, any deterioration in standards as measured by the project could be used to support calls for a "back-to-basics" curriculum, for more standardisation, more central control and greater teacher efficiency.

PROFESSIONAL ASSOCIATIONS

Included here were groups such as English and mathematics subject associations of teachers, State Institutes of Educational Research, Chapters of the Australian College of Education, Reading Associations, and so on. The most dominant feature of the responses was the diversity of opinions expressed. Some responses favoured national assessment, gave reasons for this view and argued a case for using results for monitoring purposes. Other responses indicated uncertainty and gave arguments for and against the project. Yet others expressed opposition to the project and argued that it could undo many positive things being done in schools.

The diversity of opinions occurred not only across groups but also within groups and within States. For example, in one State, the mathematical association argued against the project while the association for the teaching of English argued in favour. (Responses were completed by the respective association Presidents). Similarly, in some States, the Institutions for Educational Research were in favour while in others, opposition was expressed. In one State, the

Institute took the trouble to survey a sample of members for their opinions on aspects of ASSP. The views expressed by members within this one group similarly ranged across the whole spectrum.

EMPLOYER GROUPS AND INDUSTRIAL AGENCIES

State Employers' Federations and Chambers of Commerce and Industry expressed strong support for national assessment of literacy and numeracy. These groups saw a need for public accountability and comparable "standards" and saw national assessment as providing appropriate checks on these. Most responses indicated that little or no change would be expected since 1975, but some also predicted that standards of writing and numeration would be found to be generally unsatisfactory. Most saw positive benefits accruing from the testing although these were expressed in fairly general terms. However, one State Chamber of Commerce and Industry predicted that the ASSP findings would result in a curtailing of Federal spending, a critical appraisal of teaching, greater community demands on teachers and more emphasis on basic skills in curricula. While views of this kind were not widespread, they form part of the context which has led to the stand taken by teachers' unions.

Questionnaires were returned by Trades and Labor Councils in two States. Both initially expressed support for the project in general terms. However, one of these was subsequently modified to accord with findings of the ACTU Educational Conference of September, 1980. The revised opinion indicated that the survey was unlikely to reveal anything not already known and that the money would better be spent if it were committed to solving practical problems in education.

The Australian Public Service Board also responded to the questionnaire. The response indicated that the Board, as a major employer, was interested in levels of literacy and numeracy amongst school leavers. On balance, the Board favoured national assessment but acknowledged the dangers inherent in misinterpretation of results. The view was expressed that long term benefits could result if particular problems were identified and acted upon. The Board wished to consider the findings in deciding how to weight its own tests and school assessments when recruiting staff.

One State Department of Employment and Youth Affairs favoured the project and expressed the hope that aspects of the final report would assist those responsible for planning the Commonwealth School to Work Programs and the Educational Programs for Unemployed Youth.

ADMINISTRATORS IN NON-GOVERNMENT SCHOOLS

Questionnaires were forwarded to administrators of various systemic schools (for example, Catholic, Seventh Day Adventists), associations of independent schools, associations of headmasters from independent schools, associations of heads of independent girls' schools, and so on. The views expressed by directors in Catholic systems tended to vary from State to State with some favouring national assessment and others sceptical about its value. Perhaps the most common view expressed by the other groups listed above was that

the association had no clear policy on the matter and preferred independent schools to speak for themselves. Several responses indicated a desire to know more of the purposes and functions of ASSP. Major non-government school systems had agreed to participation in the project following requests from Directors-General for their co-operation.

POLITICAL PARTIES

Only a limited number of responses were received from spokespersons for the major political parties in the various States. All responses indicated support for national testing in some form, feeling this gave recognition to a serious problem and would provide a useful monitoring of achievement in basic skills. However, one Shadow Minister referred to the adverse comments he had received from parent and teacher groups concerned with potential implications associated with perceiving students, and indirectly teachers, in competition with each other.

OTHER GROUPS

Opinions were also sought from a number of other groups not readily classified in the categories above. Some were specialist groups (for example, a service centre for tertiary institutions) while others drew on a wider community (for example, responses were received from Country Women's Associations in two States). Responses from these other groups are not reported here but they did help to give a more complete picture of expectations held and this is reflected in the summary of opinions given in the following section.

SUMMARY OF CONCERNS AND EXPECTATIONS

Naturally it is difficult to provide a succinct summary from a set of responses which varied so much, both in terms of understanding of the project and in terms of expectations for what it might achieve. However, it is significant that opinions on the value of the project were sharply divided, even within similar goups.

In general, teachers' unions and, to a lesser extent, parent groups indicated most opposition to the project and indicated that ASSP was not really geared to provide assistance to students and teachers who needed assistance most. Hence these groups argued for other forms of evaluation which could be followed by direct remediation. They drew attention to broader curriculum objectives and warned against misinterpretation of results. The arguments of these groups, whether for or against national assessment, were couched in specific terms and usually related closely to classroom practice. Accountability was seen as a function of the professional role of individual teachers.

At another level, a numer of professional associations were clearly divided in their attitudes to, and expectations of, the project. Balanced arguments were presented but conclusions varied. While some saw value in national monitoring, others were sceptical. The arguments raised here were usually more general and tended to

relate more to curriculum issues and to system-level planning than specifically to classroom practice.

Finally, at the opposite end of the continuum, employer groups and some community groups tended to favour national assessment. The arguments here were the most general and usually referred to "standards", to the need for system accountability and the desirability of monitoring to achieve it.

CONCLUSION

This chapter first examined what can be theoretically expected of ASSP, given the decisions which were taken about the content to be tested, the nature of the tests, the defined standards of acceptable performance, the students to be involved, additional data to be collected and the mode of reporting. Later sections of the chapter reported on the community's actual concerns and expectations of ASSP, the latter being based on questionnaire responses from about 60 organisations.

Two major themes have emerged from our analyses. The first is the widespread lack of appreciation in the community of the limited nature of the 1980 ASSP project and, consequently, of implications which can validly be drawn in terms of what the project might be expected to achieve. In the main, however, we are here dealing with matters of fact rather than opinion, so that much of the misunderstanding and confusion surrounding what are perceived to be potential outcomes of the testing program can be dispelled through the provision of adequate, accurate information.

The second theme revealed in this chapter relates to the great diversity of views in the community at large and within professional educator circles in particular, on the desirability of a national assessment program, even one conducted on a limited basis. Here, however, we are confronted with matters of opinion, not of fact. That is to say, people adopt different value stances towards national assessment which are ultimately based on their beliefs about what the aims of schooling should be and what directions schooling should take. Value stances may be related to roles. For example, a Minister may view the information provided by ASSP more highly than does his Director-General, and the views of teachers and parents may reflect their particular role in the educational process. Such differences are translated into different educational priorities, and these in turn determine how, in the opinion of the individual or group in question, available resources of time and money should be allocated.

It is abundantly clear that no testing program such as the 1980 ASSP project can resolve value-stance differences of this nature. Policy makers, however, do need to take a decision on the future of the ASSP project from their assessment of the benefits against the cost and in the light of competing alternatives. Given divided opinions from varying persons and groups, it will be necessary to explain carefully to an interested community, not only the nature of the decision, but also the reason for it. Not to do so appears likely to produce dissension and conflict which can only be detrimental to the education of our young people.

REFERENCES

1. AEC press release, Australian Studies in Student Performance, July, 1980.

2. Bourke, S.F. and Keeves, J.P. (eds.) Australian Studies in School Performance Vol.III Canberra: AGPS, 1977, pp.8-9.

3. Burton, N.B. Societal standards. Journal of Educational Measurement, 1978, 15, 263-271.

4. Opinions from Ministers of Education and Directors General were sought separately through interview, and reported in Chapter 1.

5. While this was a convenient grouping for our purposes, of course we would not wish to imply that a number of groups listed elsewhere were not professional.

CHAPTER FOUR

THE TESTING PROGRAM

Barry McGaw and Donald Hogben

INTRODUCTION

The connection between the 1975 and 1980 testing programs conducted by ACER was important both for the development of the tests and for the analysis of the results. The 1975 tests were used in a survey of literacy and numeracy in Australian schools for the House of Representatives Select Committee on Specific Learning Difficulties. The 1980 tests were used to monitor levels of literacy and numeracy for the Australian Education Council but also to provide comparisons with 1975 levels.

For both surveys, functional skills in literacy and numeracy were conceived in terms suggested by a 1963 UNESCO definition.

> A person is literate when he has acquired the essential knowledge and skills which enable him to engage in all those activities in which literacy is required for effective functioning in his group and community, and whose attainment in reading, writing and arithmetic make it possible for him to continue to use these skills towards his own and the community's development and for active participation in the life of his country (1,2).

For 10 year-olds, 'effective functioning' was defined in terms of the requirements of further schooling. For 14 year-olds, at or close to the end of compulsory schooling, 'effective functioning' was defined in terms of the requirements of adult life in the Australian community.

The use of a common notion of functional literacy and numeracy in the 1975 and 1980 surveys and the desire to compare performance levels on the two occasions resulted in the use of some common items in the 1975 and 1980 versions of each of the tests.

DEFINING MINIMAL COMPETENCE

Central to the testing in both the 1975 and the 1980 surveys was an attempt to define levels of minimal competence in the domains of reading, writing and numeration. Test items were developed to discriminate between those who could perform at or above the defined

levels of minimal competence and those who could not. Criteria for successful minimal performance were not specified for each item. Rather, for collections of items comprising a sub-test, successful performance was defined to be correct answers on 80 per cent or more of the items.

Tests of this type, "constructed so as to yield measurements that are directly interpretable in terms of specified performance standards" (3), are described as criterion referenced tests. In the report of the 1980 survey, the authors summarize the classic arguments in favour of criterion referenced measurement without giving any account of the recent debate about its weaknesses (4). In particular, no attention is given to the arbitrariness of the manner in which criteria are established.

The current debate is not about the role of judgment in defining a test domain, setting standards and developing items. It is recognized that judgment is inevitably involved in any test development exercise, even one designed to produce tests which compare individuals with one another rather than with some external performance standard. What is more contentious is the belief that the external standard for criterion referenced measurement can be established in a non-arbitrary way.

Competence in any domain can be thought of as a continuum "stretching from 'absence of the skill' to 'conspicuous excellence'" (5). Determining the point on the continuum to be considered as 'minimal competence' is at best a task of arbitrary definition unless some attempt is made to validate the judgment. Scriven (6) acknowledges the initial arbitrariness of any attempt to define and test minimal competence but urges, not abandonment of the effort but follow-up to determine how well the test-based categorization into 'competent' or 'not competent' predicts later performance levels.

ACER can be excused the initial use of arbitrary measures of minimal competence in 1975 in the first national attempt to obtain any indicators of performance levels on basic skills. (Full details of the principles and procedures used in the testing program were reported after the 1975 study). The problem is that this initial survey is now enshrined as the benchmark against which all subsequent assessments can be compared. The political rhetoric discussed in Chapter Two reflects no doubts about the validity of the conclusions drawn in 1975.

ACER alone, however, cannot be expected to bear the blame for repeating the exercise in 1980 with the same arbitrariness. This could have been avoided if attention in the meantime had been given to validation of the 1975 testing, but who was there to commission such research? Funds have been available to establish the certainties, not to explore the doubts. We still do not know how well the classification of 10 year-olds as 'competent' or 'incompetent' in basic skills predicts their further school performance. We certainly do not know how well the classification of 14 year-olds as 'at least minimally competent' or 'incompetent' predicts their performance in adult life.

Burton (7) actually rejects the notion of minimal competence as a basis for establishing performance standards. She points out that real life successes have many potential causes.

> No single skill is so essential that it can be defined as necessary for survival. There may be some 'basics' that are necessary to survival but they are of the order of life, liberty, food and shelter. There is another set of basics that is quite inconvenient to do without - such as hearing and, to a lesser extent, reading. But even these most basic skills can do no more than affect the probability that one will have a comfortable life.

Burton thus claims that the complexity of causation of success in adult life renders useless the notion of minimal competence. Glass (8) does not abandon the competence continuum but claims that the location of a level of minimal performance on it is inevitably and continuously arbitrary. Scriven (9) accepts that the location of the minimal level is arbitrary initially but claims that validation studies can make it less so.

What the Australian experience has now provided is an initial arbitrary definition and assessment of minimal competence, enshrined as a baseline assessment without regard to its arbitrariness, and a repetition of the assessment six years later without any validation of the original judgment in the meantime.

MINIMAL COMPETENCE IN READING

The central component of literacy is taken to be reading for which the definition accepted by ACER is 'the acquisition of meaning from print, or ..thinking stimulated by print' or, alternatively 'asking questions of a printed text' (10). Three particular sub-skills were used to define the domain further and to guide test development. They were 'decoding and related word-attack skills', 'comprehesion' and 'practical literacy, in the sense of a person's ability to use written...information to achieve practical goals' (11).

The use of these three sub-skills highlights the problem of definition of minimum competence. If the first skill is a pre-requisite for the other two it is, in fact, more minimal than they are. The requirements of adult reading in our society can be claimed to require the interpretation of continuous prose, the assignment of an appropriate meaning to a written text. Adequate performance on decoding tasks, even adequate performance in comprehending simple sentences, may be no guarantee of adequate performance in comprehending extended prose. Whether they are is, of course, an empirical question but one that is ignored if all sub-skills just become part of the definition of functional literacy. To admit the pre-requisite skills to the definition of minimal competence is to lower the standard. To include them in the test at least allows an investigation of the empirical question. Such is the arbitrariness of definition!

MINIMAL COMPETENCE IN WRITING

The discussion of the development of the writing tests provides no precise statements or lists of objectives which make clear the test developers' understanding of writing as an aspect of literacy. In the section 'Identification of Objectives' (12), however, writing is described as "a purposeful activity in which several systems of knowledge or skills are integrated in the responses which are made to writing demands". Because the focus is on the production of a response through an integrated use of skills, the definition does not include specific isolated knowledge, such as knowledge of grammar or word usage, or isolated skills such as handwriting, spelling, punctuation and capitalization.

MINIMAL COMPETENCE IN NUMERATION

As with the discussion of writing, the discussion of numeration provides no statement of the precise meaning of numeracy. It is declared that the numeration tests were not intended to 'cover all, or even most, topics in the mathematics syllabuses of schools, but only those aspects related to tasks considered necessary to either school or community requirements for numeracy' (13).

Seven content areas within the mathematics curriculum were identified as important for numeracy but emphasis was given to arithmetic computation and measurement. In an appendix to a separately published description of the development of the 1980 numeration tests, Bourke (14) lists 177 specific competences within numeracy. A priority rating from 1 to 3 is given to each competence though no indication is given about the criteria used in establishing these priority ratings.

Three levels of ability, knowledge, computation and application, were also identified as important for tests of numeration. These abilities were assumed to be hierarchically related which again raises the question of how a general definition of 'minimal competence' can be arrived at when it is conceived in terms of sets of skills, some of which are pre-requisite for others in the set.

A further difficulty arises from the failure to define numeracy other than by listing mathematics content areas and levels of ability. It leaves the concept loosely defined and subject to arbitrary variation according to the emphases assigned to the components. If a more general definition such as 'the ability to represent and read mathematical concepts written in mathematical symbols' were used it would be more like the definition of literacy developed in terms of reading and writing. Within such a definition, computation would probably not achieve the status accorded it in the numeration tests. But then it may be that computational skill is what is really meant and wanted. If that is the case, the survey could be said to be of literacy and computation rather than of literacy and numeracy.

MEASURING MINIMAL COMPETENCE

Development of the Tests

Thirty minute tests of reading, writing and numeration were developed for both 10 year-old and 14 year-old levels. All were approved by the project's Steering Committee, representative of all State and Territory Education Authorities, after special scrutiny by its Technical Sub-committee. Melbourne based Development Committees were used for each test and draft versions were circulated to Education authorities represented on the Steering Committee. Unfortunately the substance and impact of criticisms from Curriculum specialists in State Departments is unknown. All trial tests were field-tested in versions which contained sufficient items from the 1975 tests to check that new items had difficulties comparable with those in the 1975 tests.

The test development process was elaborate and careful. The experts whose judgments of trial versions were obtained were apparently able to influence the final versions. No detail is available, however, on the nature of their advice or the extent of any changes made in the tests as a result of it. In evaluating the items, these experts had to consider not only technical adequacy of the items (in terms of clarity of wording and illustrations, layout, validity of content, etc.) but also their suitability in assessing possession of minimal skills considered relevant for all at the 10 or 14 year-old level.

The end result of this development was a set of six tests which are well produced and contain items which appear to be consistent with the developers' intentions. Since an important purpose was to compare 1975 and 1980 performance levels, several items from the 1975 tests were included in the 1980 tests. The extent of this repetition of items is summarised in Table 1. In the case of the writing tests, the proportions are expressed in terms of the time to be committed to the components rather than the number of questions.

TABLE 1: PROPORTION OF 1980 TESTS CONSISTING OF ITEMS FROM 1975 TESTS

	10 year-old	14 year-old
Reading	6/27	7/35
Writing	9/30	15/35
Numeration	13/35	13/35

While repetition of some test content like this is essential if results from separate testings are to be compared, there can be difficulties. Schools may actually practise students on the previous version of the tests and thus actually practise them on some of the items in the new version. There is probably little likelihood that this happened for the 1980 testing. The 1975 tests were available in various reports but not readily accessible to schools. In addition, the evidence cited elsewhere in this report suggests that schools paid little attention to the 1980 testing beforehand although the general information sheet provided to schools by ACER made explicit reference to the 1975 testing and to the intention to compare 1975 and 1980 performances (15). If repeated monitoring were to become a regular feature of the educational scene then the problem of practice effects would be likely to be more substantial.

The reading tests used for 10 and 14 year-olds are reproduced in Mills and Holzer's report (16) on the development of those tests. The test for 10 year-olds contains 27 items, but only 8 of them deal with comprehension of continuous prose, 3 of those relating to newspaper articles. Comprehension of single sentences is required for 10 items while the remaining 9 test the use of reference skills with tables of contents, indexes and so on.

The comprehension tasks included in the tests varied significantly in difficulty. There were 5 simple sentences which required selection of a matching illustration from among four and they were predictably very easy. The final performance rates confirm this, with 99 per cent of 10 year-olds in the sample being successful. There were 5 further sentences on a more complex form for which the questions demanded that inferences be drawn. With the sentences having no context, these tasks were artificial and difficult. Performance on continuous prose was still lower (17). A similar pattern occured with the 14 year-olds though their levels of response were superior (18).

The point of these comments is not to offer an assessment of the levels of performance but to show how the performance data reveal the varying difficulty levels of sub-sections of the tests and the arbitrariness of these difficulty levels. With different sentences different results could well have emerged.

The writing tests are reproduced in Stanyon's report (19) on the development of the tests. Two forms of the test were produced for both 10 year-old and 14 year-old samples. These tests called for writing of different types (letters, messages, stories, etc.) for a range of real-life settings. There was an ambiguity in two questions which may have distorted some results. In one of the 14 year-old tests, the students were given a map showing the location of a school and a library and were told to write directions for someone wanting to drive from the school ground to the library. It was not made clear whether the person for whom the message is to be written would have access to the map. In one of the 10 year-old tests, a similar problem could have occurred with a "write a story" task based on an illustration if the 10 year-olds were unsure whether their readers would also have the illustration. ACER reports that there was no

evidence in the student responses that these potential ambiguities caused any problems.

The marking schemes developed for assessing writing performance are clearly outlined in the report as they were given to the teacher-markers. Evidence provided of reliability checks on markers suggests the strategies adopted were adequate. Evidence is also provided of the number of words and spelling error rates in performances of the students over the various writing tasks. In any test of this type it is difficult to control for the effects of background, knowledge, interest and motivation on the performances on specific tasks. The more realistic the task, such as the 'writing directions' task, perhaps the less influential these factors will be.

The numeration tests are reproduced in Bourke's report (20) on the development of the tests. A significant feature of these tests was the use of free response and the limited reliance on the multiple choice format. The test items are well presented and clear, with a heavy emphasis on arithmetic computation as indicated earlier and as acknowledged by Bourke. Some items appear to test at levels other than indicated. For example, declaring that the 5 in 3251 stands for 5 tens is probably more than just a knowledge item. Given the way this knowledge is taught before Year 5, ACER claims that the classification is probably correct.

The few reservations expressed above about particular items on the tests are somewhat gratuitous. The overall picture of the tests is one of a set of instruments carefully and appropriately prepared for the task for which they were intended.

ESTABLISHMENT OF CUT-OFFS

Glaser (21) developed the notion of "criterion referenced measures" to emphasise the importance of the content of a test for the interpretation of test scores. Using a notion similar to that of the traditional "content validity", he emphasised that the test items define the criterion on which performance is assessed and suggested that the criterion should be thought of as a continuum along which levels of performance can be registered. Glass (22) claims that Glaser's notion of criterion referencing was subsequently coupled with Mager's (23) arguments for clear specification of the level of performance of some particular behaviour required to indicate achievement of some instructional objective. According to Glass, this coupling, by writers such as Popham and Husek (24), led to the criterion being understood as a particular required level of performance rather than a continuum on which any level of performance could be assessed. The measurement task became one of defining the required performance level rather than the continuum.

The level at which the required performance is set depends on the purpose of the measurement. If selection for admission to some restricted activity is being undertaken, it can be set appropriately high. If certification that some minimal performance level has been achieved it can be set appropriately low. For the ASSP, the

assessment task has been constructed as the assessment of whether students are above or below defined minimal levels of numeracy and literacy. The measurement task is to define single points on presumed continua from illiteracy to literacy and from innumeracy to numeracy.

Specific test items can be arranged along a continuum such as that from illiteracy to literacy. Items which can be answered with little literacy competence are easy items located at points towards the "illiterate" end of the continuum. Students who are unable to answer them correctly are further down towards illiteracy. Students who can answer them correctly are further up towards literacy.

When the measurement task is concerned with classifying persons as above or below some point, there are two tasks. The first it to define the relevant point on the continuum. The second is to develop test items which are located at that point. The success or failure of individuals on those items then indicates whether their performance is above or below the cut-off.

For the 1975 and 1980 surveys, ACER defined the cut-off points for minimal competence in the manner already described. In fact, more than one point was defined on the various continua. With reading, for example, comprehension of simple sentences, complex sentences and continuous prose represent different points on the literacy continuum since they are tasks of differing difficulty requiring different levels of skills for successful performance. It does not matter that three points are defined and that performances are judged with respect to each. It does, however, make clear that the total test scores of individuals provide little information about levels of mastery. The message of the surveys lies in the sub-test scores.

The surveys thus do not measure performance with respect to a single criterion point but rather with respect to several distinct points. The points are, however, relatively close together on the overall continuum in a range well below the performance levels of the great majority of the students.

An alternative to focusing on this small range of the continuum in the minimum competence region would be to broaden the concern of the testing from a classification into competent and incompetent to an assessment of the distribution of persons along the whole continuum. The tests would need to be longer and the samples larger, but the information would be much richer. It would still be possible to define points of particular concern and to estimate proportions of the population above and below them. The needs for monitoring minimal performance could thus still be met. The total data set would, however, provide an educationally more useful picture of the total range of skill development in numeracy and literacy. An important risk is that the testing program could have a much greater controlling influence on what is taught in schools.

Focusing on the full continuum has further advantages in the development of new items. There is less need to judge, *a priori*, the exact location of an item on the continuum. As long as items are generally spread along it, they can provide a basis for estimating the

location of each student on the continuum.

When competence is defined as success on 80 per cent of the items in a sub-test, it is much more important that all items in the sub-test are at the same point, that is have the same difficulty level. Replacement of an item by one marginally more difficult would lead to lower estimates of overall performance; replacement by one marginally easier would lead to higher estimates. Focusing on the continuum rather than on particular points removes the necessity of a strong assumption of sub-sets of items having equal difficulty. The separate locations of all items can be estimated from the responses together with the locations of individuals. An analysis of the existing data sets in these terms, even for the short ranges of the continua involved, would be worth trying using the Rasch Simple Logistic Model (25). Such work is being undertaken at ACER.

In the 1975 ACER survey, mastery level on a sub-set of items was defined to be a level of performance which produced correct answers 80 per cent of the time. This method was used again in the 1980 survey. Cutting scores on particular sub-tests were calculated by a Bayesian method outlined in the report (26) but described in detail by Morgan (27). In arriving at the cutting score, a loss ratio of 1:1 is used to treat as equally significant misclassifications of competent persons as incompetent and misclassifications of incompetent persons as competent.

ESTABLISHING NATIONAL ESTIMATES

Sample Size

To obtain national estimates of performance levels it would have been most efficient to select a simple random sample with a uniform sampling fraction obtained from all States and Territories. The need to provide estimates of performance for State or Territory sub-populations, however, meant that larger samples than would otherwise have been necessary were drawn from the smaller States and the Territories.

The use of two-stage stratified random sampling (sampling schools and then students within schools) was the only sensible way to obtain an accessible sample of the size required. More students were tested than would have been required in a simple random sample but many fewer schools were involved. The planned two-stage sample of 7000 would have been equivalent to a simple random national sample of 1030.

For the sampling, 40 schools in each State and 20 schools in each Territory were originally selected for both the 10 and 14 year old samples. The schools were selected from State and Territory lists arranged separately in postcode order for government and non-government schools.

If a school declined to participate, the next school on the list was invited. Within each school 25 students of the appropriate age were selected at random by ACER for the testing. If a school had

fewer than 25 students available, the additional students were selected from the next school on the list. Replacement students were also selected at random and held as reserves to be tested in place of any students in the original sample who were unavailable at the time of testing.

These strategies for replacing schools or students are based on the assumption that each school is equivalent to the one which follows it on the postcode order list and that, if any students are unavailable, those drawn from the replacement list will be equivalent to them. All students included in the sample as replacements were identifiable. The crucial consideration in deciding whether their inclusion biased the final estimates in any way is not whether they differ from the other students in the sample in any significant way but whether they differ from the students they replaced. The report of the survey demonstrates that the replacement students were not significantly different from other students in the sample in any of the domains tested. No data have been gathered on their similarity with the students they actually replaced. It would be worth undertaking a small study to determine whether any schools managed to have poorer students systematically replaced by better ones.

The use of equal sample sizes from all States and equal sample sizes from the two Territories, instead of a common sampling proportion from each of these sub-populations, necessitates the weighting of estimates for these sub-populations in the computation of estimates for the national age population.

The relevant weights can be defined by a formula such as that given by Bourke (28). That particular formula is designed to make further adjustments because of the differentially low response rate in some sub-samples. Had the designed samples been achieved such a formula would still have had to be used. The precise function of the formula might be clearer if it were expressed as

$$w = \frac{\text{Size of Australian age sample}}{\text{Size of Australian age population}} \times \frac{\text{Size of sub-population}}{\text{Size of sub-population sample}}$$

The left hand term in this expression is a constant for all sub-populations at a particular age level. For the designed samples it would have been,

$$\text{for 10 year olds} : \frac{7\,000}{259\,705} = .0270, \text{ and}$$

$$\text{for 14 year olds} : \frac{7\,000}{242\,697} = .0288$$

The purpose of this constant term is to scale the weights to the total sample size of 7,000. This is misleading since it tends to convey the impression that the accuracy of the sample is of a higher order than it really is. It would be more appropriate to scale the weights to the size of the equivalent random sample of 1,030 or the relevant Australian age population instead. If the latter were chosen the scaling term in the expression for the weight would be eliminated.

The right hand term in the expression for the weights simply adjusts the contribution of a State or Territory's estimates to the Australian estimates to make it proportional to the age sub-population and not the planned sample size for the State or Territory. The unscaled weights which would have been used had the designed sample been achieved are those shown in Table 2.

TABLE 2: UNSCALED WEIGHTS FOR COMBINING SUB-POPULATION ESTIMATES TO OBTAIN POPULATION ESTIMATES

Sub-population	10 Yr Old Sample	14 Yr Old Sample
ACT	4327/500 = 8.654	3847/500 = 7.694
NSW	86409/1000 = 86.409	81533/1000 = 81.533
VIC	71379/1000 = 71.379	66147/1000 = 66.147
QLD	40229/1000 = 40.229	37882/1000 = 37.882
SA	22964/1000 = 22.964	22646/1000 = 22.646
WA	23703/1000 = 23.703	21495/1000 = 21.495
TAS	8160/1000 = 8.160	7575/1000 = 7.575
NT	2534/500 = 5.068	1572/500 = 3.144

Response Rates

In the 1980 study, considerable problems of non-response were encountered at a late stage in the preparations for the survey. A substantial proportion of the schools which had already indicated a willingness to co-operate withdrew from the study as a result of pressure from teachers' unions and parent organizations.

This withdrawal rate was uneven across States and Territories as Table 3.2 of the report indicates (29). It was also uneven between government and non-government schools. Although no explicit response rates are given for type of school, it may readily be inferred from Table 3.3 (30) that non-response was particularly severe among government primary and secondary schools in the ACT, Queensland and Victoria and government primary schools in Tasmania.

Unequal sampling proportions were planned in the different States and Territories with the intention of using weight adjustments proportional to those in Table 2 to correct for the different levels of representation. To adjust for the non-responses, ACER simply changed the unscaled weights from those in Table 2 by using the size of the achieved sub-sample instead of the size of the designed sub-sample in the computation of the weights. Since the response rates were different for government and non-government schools in most of the States and Territories, separate weights were calculated for all these sub-samples. The scaled weights calculated by ACER are shown in Table 3.3 of the report. In Table 3 these are shown as unscaled weights, together with the unscaled weights which would have been used had the designed sample been achieved.

TABLE 3: UNSCALED WEIGHTS FOR USE WITH DESIGNED AND ACHIEVED SUB-POPULATION ESTIMATES

Sub-population	10 Yr Old Sample			14 Yr Old Sample		
	Designed	Achieved		Designed	Achieved	
		Gvt.	Non-gvt		Gvt.	Non-gvt
ACT	8.7	25.2	6.4	7.7	114.8	8.1
NSW	86.4	89.7	91.5	81.5	85.4	84.3
VIC	71.4	106.3	70.6	66.1	131.6	64.9
QLD	40.2	71.8	38.2	37.9	90.4	37.6
SA	23.0	24.2	24.0	22.6	22.7	23.9
WA	22.7	25.0	23.8	21.5	25.3	22.4
TAS	8.2	21.7	8.3	7.6	12.4	7.9
NT	5.1	6.1	4.8	3.1	7.5	3.3

In all cases, the response rates for non-government schools were very high so the weights required for the achieved sample were virtually the same as those which would have been required for the designed sample. Similarly high response rates for government schools in NSW, SA and WA also resulted in weights little different from those for the designed sample.

In ACT, Tasmania and NT the non-responses reduced the size of the achieved sub-samples towards levels which were not so uneconomically high, from the point of view of contribution to the estimation of national performance levels. The weights for these sub-samples were increased to levels closer to those for NSW. In the case of ACT non-government secondary schools, the 14 year old sub-sample was so small that its weight of 114.8 actually exceeded that for NSW but was clearly closer to the NSW level than the original 7.7. Even for Queensland, a more populous State, the non-response rate did not raise the weights to levels substantially above those for NSW. The nett effect of the sample reductions in all these cases was to make the sample more efficient.

Only in Victorian government schools, and particularly at the secondary level, was non-response severe enough for differential weighting to reduce sample efficiency.

The weighting of sub-sample estimates summarized in Table 3 takes account only of the reductions in the size of the samples. It provides no control of bias in the sub-sample estimates of the type which would occur if the non-responding schools differed systematically from the responding schools.

Adequacy of Estimates

The relatively high non-response rate in some of the sub-samples does not appear to have biased the national estimates. The fact that there was a high response rate in NSW, the most populous State, gave stability to the national estimates. It also provided a basis for comparisons of inter-State differences in 1975 and 1980 which suggested that the low 1980 response rates had made little difference. The comparisons of Victoria and Queensland with NSW (31) provide fairly strong evidence that serious bias was not introduced into the Victorian data, in comparison with NSW.

A potentially more significant source of bias could have been the replacement strategy for students unavailable for testing. Absentees can be expected to include a relatively large proportion of underachievers. The same might be true of the group unavailable for testing because of "the organizational pattern of most secondary schools where students come together in different groups for different subjects" (32). There may have been a tendency for students in remedial classes, for example, to be withheld from testing. Some comparison of replaced and replacement students in terms of their level of school performance would have been valuable.

The overall estimates of performance levels may have been elevated slightly by the absence of this sub-set of the original sample and the presence of the replacement students. Provided the extent of replacement was similar for the 1975 study, comparisons of performance levels in 1975 and 1980 would be unaffected.

CONCLUSION

Any decision to use a national sample to monitor performance at levels of minimal competence is contentious both politically and technically. In this chapter the technically contentious issues have been examined.

The essential arbitrariness of definitions of minimal competence was seen as a significant problem. The task of designing tests with items with difficulty levels all focused at the point of minimal competence was seen to be substantial and, in addition, likely to give much less valuable returns for funds invested than the alternative of preparing tests which assess performance over a wider range of difficulties along a continuum of increasingly good performance.

Given that ACER was committed to assessment at the level of minimal competence, their task can be judged to have been well performed technically. The tests were good and the national estimates of performance acceptably unbiased, the latter despite very uneven response rates across the country.

REFERENCES

1. Keeves, J.P. and Bourke, S.F. Literacy and Numeracy in Australian Schools: A First Report. Australian Studies in School Performance Volume 1. ERDC Report No.8. Canberra: AGPS, 1976, p.6.

2. Bourke, S.F., Mills, J.M., Stanyon, J. and Holzer, F. Performance in Literacy and Numeracy: 1980. Melbourne: Australian Education Council, 1981, p.10.

3. Glaser, R. and Nitko, A.J. Measurement in Learning and Instruction. In Thorndike, R.L. (Ed.) Educational Measurement Washington: American Council on Education, 1971, p.653.

4. Bourke, et al., op. cit., pp.7-10.

5. Glass, G.V. Standards and Criteria. Journal of Educational Measurement, 1978, 15, 237-261.

6. Scriven, M. How to Anchor Standards. Journal of Educational Measurement, 1978, 15, 273-275.

7. Burton, N. Societal Standards. Journal of Educational Measurement, 1978, 15, 263-271.

8. Glass, op. cit.

9. Scriven, op. cit.

10. Bourke, et al., p.11.

11. Ibid.

12. Ibid, p.18.

13. Ibid, p.24.

14. Bourke, S.F. Development of the Numeration Tests, 1980: A Statement of Objectives and Related Tasks. Melbourne: ACER,

1981.

15. Bourke, et al., op. cit., pp.173-4.

16. Mills, J.M. and Holzer, F. Development of the Reading Tests, 1980: A Statement of Objectives and Related Tasks. Melbourne: ACER, 1981.

17. Bourke, et al., op. cit., pp.69-73.

18. Ibid., pp.73-78.

19. Stanyon, J. Development of the Writing Tests, 1980: A Statement of Objectives and Related Tasks. Melbourne: ACER, 1981.

20. Bourke, op. cit.

21. Glaser, R. Instructional Technology and the Measurement of Learning. American Psychologist, 1963, 18, 519-521.

22. Op. cit., pp.240-241.

23. Mager, R.F. Preparing Instructional Objectives. Palo Alto, California: Feardon, 1962.

24. Popham, W.J. and Husek, T.R. Implications of Criterion-referenced Measurement. Journal of Educational Measurement, 1969,6, 1-9.

25. Wright, B.D. and Stone, M.H. Best Test Design. Chicago: MESA 1979.

26. Bourke,et al., pp.52-57.

27. Morgan,G. A Criterion-referenced Measurement Model with Corrections for Guessing and Carelessness. Occasional Paper No.13. Melbourne: ACER,1979.

28. Bourke, et al., op. cit., p.38.

29. Ibid, p.37.

30. Ibid, p.39.

31. Bourke, et al., op. cit., pp.42-44.

32. Ibid., pp.64-65.

CHAPTER FIVE

IMPLEMENTATION OF THE TESTING PROGRAM

John Theobald and Colin Power

INTRODUCTION

The schools participating in the project were promised confidentiality of identity with respect to their participation in the project by the AEC and ACER. The evaluation team did not have access, therefore, to information regarding the identity of schools being approached by the ACER. Clearly if the evaluation team was to study the impact of the program on participating schools, at least some of the schools involved had to be studied. The ACER therefore agreed to select another sample for the evaluation using the same procedures as had been used in selecting schools for the testing program. The ACER selected a sample of 526 schools, half of which had been asked to participate in the 1980 ASSP testing and half of which had not. However, no information on whether or not an individual school had been approached to participate in the testing was provided,

The schools in this sample were approached late in September 1980, to determine whether:

(a) They had been asked by ACER to participate, and

(b) If so, whether they would be willing to participate in the evaluation by completing questionnaires and/or by having a member of the team visit the school.

This approach took place immediately prior to the October testing and took the form of a letter about the project together with a brief questionnaire and a reply paid envelope. Schools were again promised confidentiality of identity in accordance with AEC requirements. Immediately after testing, schools which had agreed to participate in the evaluation were visited or asked to complete a questionnaire on the implementation of the testing program.

The responses to the questions asked by the evaluation team provide information regarding the sample utilized in the evaluation and aspects of the implementation of the project as seen by the participating schools.

CHARACTERISTICS OF THE EVALUATION SAMPLE

The sample of schools provided by ACER included 40 primary and 40 secondary schools from each state (except Tasmania where 23 secondary schools were selected), 20 primary and 11 secondary schools from the ACT and 20 primary and 12 secondary schools from the N.T.

The response rates from schools to the initial approach by the evaluation team are shown in Table 1.

TABLE 1: Response rates to initial request, by state

State or Territory	Schools Approached	Schools Responding	% Participation
N.S.W.	80	51	64
VIC	80	64	80
QLD	80	59	74
S.A.	80	70	88
W.A.	80	66	83
TAS	63	45	71
ACT	31	24	77
N.T.	32	31	97
TOTAL	526	410	78

Of the schools responding 81 per cent were government, 14 per cent Catholic and 5 per cent Independent; 56 per cent were located in a metropolitan area, 20 per cent in a provincial city (population > 10,000) and 24 per cent in a rural area; 51 per cent were primary and 49 per cent secondary schools; 4 per cent had enrolments under 200, 48 per cent enrolments between 201 and 600; and 38 per cent enrolments over 600.

AWARENESS OF ASSP PRIOR TO THE OCTOBER TESTING

Most schools (78 per cent) responding to our initial approach indicated that they were aware that a national survey of the performance of 10 and 14 year olds in aspects of reading, writing and numeration was about to be undertaken. Many (238 out of 410) had been approached to participate in the study and had learned of it through official channels. Of the remainder, 51 per cent had heard of the project, their major sources of information being the media and professional associations. Of the schools not approached to participate in the testing program, half had not heard of the national

survey prior to receiving the letter from the evaluation team.

PARTICIPATION IN THE TESTING PROGRAM

Of the 238 schools invited by ACER to participate in the testing, 80 per cent said that they had agreed to take part. Schools were asked to indicate the reasons for their decision, who was involved in reaching the decision about participation in ASSP, and how the decision was reached.

The reasons given by schools for agreeing or not agreeing to participate in the national survey are presented below.

1. An internal benefit to the school was expected (16 per cent)

Some schools agreed to participate because they expected to obtain feedback likely to facilitate diagnosis, remediation, curriculum planning, evaluation of students' progress, teaching or school programs.

2. An external benefit was expected (44 per cent)

In most cases in which schools agreed to participate, the benefits perceived extended beyond the school; generally responses so classified indicated a belief that monitoring of national standards would provide useful information about the performance of schools.

3. School felt pressured to participate (4 per cent)

In a number of states/territories, the ACER request was accompanied by a letter from the State Director-General or his equivalent. While these letters left the decision regarding participation to the schools, at times such letters were interpreted to imply a Departmental directive. In some states/territories, schools were requested to co-operate with the ACER by the Director-General or the State Minister.

4. School felt pressured not to participate (12 per cent)

Some school responses suggested that the school had rejected the invitation or withdrawn due to external pressures (notably following advice from the Teachers' Union or other groups).

5. Internal Opposition to the program (8 per cent)

Where internal opposition within the school to the testing program was indicated, responses were classified under this code. Reasons commonly given included adverse backwash effects on the curriculum, misinterpretation and political misuse of the results, and interruptions to the school program.

6. Other (16 per cent)

Responses which could not be classified easily into any of the above categories included those in which concern was expressed over a special group (eg. aboriginal children, Vietnamese children) and those in which no explicit reason was given for participation.

The pattern of reasons given for agreeing or not agreeing to participate shows that the majority of schools, if approached, agreed to participate because of some perceived external benefit.

The overall patterns of response were fairly consistent regardless of type of school, level of schooling, or state. There were, however, minor variations among the states. The effects of Teachers' Union directives were most apparent in Queensland, Tasmania and the ACT. In S.A. no directives were issued and in N.S.W. the union circular was not sent until mid October 1980 (ie. generally too late to be effective). Internal resistance to ASSP within the school was highest in Victoria (reflecting the prolonged debate on state and national testing programs) and the Northern Territory where territory-wide testing had been mooted. In WA, a Union directive was sent to schools, followed by a counter-directive from the Director-General indicating WA had agreed to co-operate in the testing program.

It is interesting to compare the results obtained by the evaluation team with the analysis of returns from the ACER. The patterns are similar. In ACT, Victoria, Queensland, Tasmania and the Northern Territory, a sizeable proportion of schools withdrew although they had agreed to participate in ASSP, either because of teacher union or parent association pressure or because of internal resistance on the part of teachers.

Of the schools in the sample invited by ACER to participate, 69 per cent indicated that the material received from ACER provided all the information needed by the school to make a decision about participation; 6 per cent wanted more information and the remainder did not respond to the item.

For the most part then, schools were satisfied that they were adequately informed by ACER. Two schools wanted a clearer statement of the objectives of ASSP; three wanted additional information about the tests to be administered and/or administration procedures; two schools reported that some of the material sent by ACER went astray. The majority of schools seeking additional information indicated that in their view, information about the intended practical uses of the results and/or a discussion of objections to the program was warranted.

Schools were asked to indicate who in the school was involved in reaching a decision about participation in ASSP, and how the decision was reached. The responses are set out in Table 2.

The responses to the question on how the decision was reached show that in most cases (173 schools) it was an executive decision by the principal acting alone or in consultation with senior staff; very rarely (18 schools) was the decision made as a result of a staff vote or consensus.

TABLE 2: PARTICIPATION IN DECISION REGARDING SCHOOL INVOLVEMENT IN ASSP

Group	Frequency	%*
Staff		
Principal	163	69
Teachers of pupils concerned	45	19
Senior staff	66	28
Whole staff	25	11
Other (counsellors etc.)	3	1
Parents		
Parents or pupils concerned	32	14
Parents organization/School Council	11	5
Students		
Students selected	11	5

* As a per cent of participating schools (n = 238)

The explanation given by one principal would seem to sum up the approach taken in many schools:

> It is not usual in this school for me, as principal, to make decisions without full participation of staff and parents. Although I was aware of a few "rumbles" against the project expressed at the state conference of primary school principals, rightly or wrongly, I regarded the survey as a minor matter to this school, could see no harm in participating and have gone ahead as a routine matter. Both teachers and parents I know would support the project and discussion time was short for other important matters.

IMPACT ON SCHOOL IMMEDIATELY PRIOR TO TESTING

Schools were asked whether there had been any impact on school policy, curriculum teaching or evaluation procedures as a consequence of the ACER invitation to participate. Only two schools of those responding answered "Yes". They commented:

> "Evaluation of sections of the school/curriculum is being planned. The effect of this exercise will be greater after we receive the results for our school and the summary report".
> "We will be far more critical in assessing future projects".

Ten of the 187 who replied "No" added "not yet" or "may do so in the future".

WILLINGNESS TO PARTICIPATE IN THE EVALUATION OF ASSP

Of the 196 schools in the sample that had agreed to participate in the ACER testing, 183 agreed to complete further questionnaires about implementation and impact of ASSP in their school (i.e. 93 per cent of schools involved in ASSP agreed to co-operate in the evaluation study).

Schools participating in the testing program were also asked if they would be willing to have members of the evaluation team visit the school to discuss the implementation and impact of ASSP with administrators, teachers, students and parents involved: 113 indicated that they would. It was intended to select, from this pool, 5 primary and 5 secondary schools from each state and the ACT/NT to serve as case studies. An attempt was made to ensure that the case study schools in each state included representatives of different types of schools (State, Catholic, Independent) and locations (Metropolitan, Rural). However, because of the small number of schools agreeing to be visited in some states (NSW, TAS) and territories (ACT, NT) this was not always possible. It needs also to be appreciated that the time constraints on the evaluation team were very severe. The team had to be formed; negotiations entered into and agreements reached; the evaluation plan completed, instruments developed and schools approached in the brief period between AEC approval (late June) and the beginning of testing (early October). This, in turn, created difficulties in negotiating access to schools in Western Australia. The WA Education Department agreed to the evaluation team's requests, but because of timing difficulties, it was not possible to undertake case studies in WA.

IMPLEMENTATION OF THE ASSP TESTING

Immediately after the testing program in early October, questionnaires were sent to those schools which had agreed to participate in the evaluation study. These questionnaires were to be completed by the Principal or Administrator of the ASSP testing program. In schools selected for case studies, the teachers of the classes from which the students were drawn filled in questionnaires

and were interviewed, the parents of the students completed questionnaires and the students were interviewed in groups.

For the reasons given, not all of these data are available from all States. This section of the report is based on the data sources shown in Table 3.

TABLE 3: SOURCES OF DATA FOR IMPLEMENTATION REPORT[a]

Sources of data	NSW	Vic	Qld	SA	WA[b]	Tas	Total
1. Principals/Administrators[c]							
From secondary schools	10[d]	9	4	13	10	6	52
From primary schools	7	10	8	11	8	3	47
2. Number of case study schools							
Secondary	3	3	4	5	-	5	20
Primary	2	4	4	5	-	3	18
3. Teachers (by quesionnaire and interview)							
From secondary schools	7	13	24	25	-	9	78
From primary schools	8	12	14	13	-	3	50
4. Parents (by questionnaire)							
From secondary schools	37	30	75	75	-	80	297
From primary schools	35	80	133	72	-	33	353
5. Pupils (by group interview)							
From secondary schools	59	49	89	83	-	81	361
From primary schools	49	82	89	81	-	52	363

Notes: (a) Data from N.T. not available.
(b) Approval for access granted too late for case studies to be carried out in 1980.
(c) In schools selected for case study these questionnaires may have been followed by interview.
(d) Includes 2 from the A.C.T.

Differences in response rates between types of schools appear to be related to the level of opposition of state teachers' and parents' associations and also to experience with political and media response to previous surveys. Whereas in some states the ASSP program was "low key", in other states it was difficult to get government secondary schools to participate in ASSP and even more so to participate in the evaluation. Of the schools represented in Table 3, 38 have been involved in case studies, and although in individual states they may not represent an acceptable stratified sample, they include rural and urban, large and small, government and non-government schools.

The analysis of the data is grouped according to the source of the information.

PRINCIPALS

As described earlier in this chapter a pre-testing questionnaire was sent to schools to obtain information about the willingness of schools to participate in the evaluation of the implementation and impact of ASSP. The information elicited by that earlier questionnaire has been reinforced by the replies to questionnaires completed by Principals (or administrators of the ASSP program within the schools) following the administration of the tests. These questionnaires were accompanied by a covering letter and a stamped addressed envelope for reply.

It is clear that communications from the ACER about the testing were received in good time, were clear and adequate with regard to the sampling, the administration of the tests, the collection and return of used and unused materials, and that there were few, if any, difficulties in complying with the instructions given. In a few schools the testing was spread over two or three days, either because of difficulties with the school program or to catch up with students in the sample who were absent on the first or second occasion.

When schools knew which students were to be in the sample, the children were commonly assembled and told that they had been selected to do tests as part of a nation-wide survey of literacy and numeracy in Australian schools. Letters containing the details of ASSP, as provided by the ACER, were given to the students to take home to their parents. The discussion with selected students was usually carried out by the administrator of the ASSP program within the school, sometimes the Principal, but more often the Deputy Principal or a senior member of staff. Often students were told that the results were confidential and some schools, in order to allay anxiety, assured students that the results would not count towards school assessment. On the other hand, some schools told their students that it was a privilege to be chosen to represent the school and that they should do their best for the school. In one school, selected students were seen individually by the Principal and in another the names of those selected were announced at the school assembly. This information was generally given to the students about a week before the testing although some were told two weeks ahead, and a few were not told until a day or so before the tests.

Most Principals felt that the students' awareness of the testing had not affected them and gave reasons such as the following: "the whole program was "low key", "the results were confidential and didn't count in school assessment", "students are used to this style of testing", "the ASSP testing fitted easily into the school program". About one-third of the Principals reported some effect: some students were said to be anxious and apprehensive, and in some schools individual students withdrew from the testing. On the other hand, some students were reported as being proud to be chosen, pleased, excited and happy to participate.

In most schools the Deputy Principal or some other senior teacher administered the testing program and most classroom teachers were not involved in any way. There were few reports of teachers being affected by the program. Of these few, some were concerned with their Union's policy of opposition, and a stop-work meeting was held in one school. Although four Principals reported concern felt by teachers over the potential misuse of results, the main negative effect was irritation at the intrusion of the testing program into normal school routine.

As expected, it was reported that 2-3 man hours were spent in administering the ASSP tests but estimates of the man hours spent in associated activities varied widely. The associated activities described were: reading directions, listing names, discussion with students, distributing letters to parents, collating materials, timetable changes and the preparation for testing, packaging and posting materials to the ACER. The range was 1-20 man hours with a mid-estimate of about 3-4 man hours, more time being spent in the larger secondary schools.

Those who answered the question about possible effects of the ASSP survey on the school were largely divided between:

"no effects", "neither beneficial nor harmful effects" (22 per cent of respondents)
and
"don't know", "wait and see", "depends on what use is made of the data" (14 per cent).

Another 24 per cent of respondents, while not strictly answering the question, said that the survey would give feedback on levels of performance which could be compared with state or national norms. Other, less frequent, responses were: "it made the school feel as though it mattered", "it has severely damaged staff relations in the school", "useful for future planning", "will confirm suspicions about the standards of students".

A similar range of beliefs was advanced about the possible effects of the survey beyond the school. Many Principals declined to comment. The most frequent responses were:

"depends on how the data is used", "don't know" (22 per cent)
"no effects" (10 per cent)
"there is a danger of the data being misinterpreted or misused

for political purposes" (16 per cent)
"the data is likely to be interpreted as reflecting on teachers" (6 per cent)
"it will give useful informaion on national levels of performance" (12 per cent)
"can be used to establish norms" (8 per cent).

Less frequent responses included: "the survey may reassure parents that standards are higher than ever", "it will give the lie to falling standards", "will provide a basis for evaluation", "will help in developing realistic programs" and "teachers may teach toward the tests".

TEACHERS

In those schools selected for case study, questionnaires were distributed to the teachers of the classes from which the students in the sample were drawn. Further detail was obtained by following up these questionnaires with interviews of the teachers concerned.

About half (46 per cent) the teachers first learned about the survey through discussion with other members of staff, and about a quarter (22 per cent) directly from the Principal. Of the remainder, 17 per cent had read the printed material distributed by the ACER, 9 per cent had read a press release, a few had become aware of the survey through union activities and two teachers claimed that they still knew nothing about the tests or their results.

Similarly, about half the teachers who responded (48 per cent) believed that the information received from the ACER adequately explained the nature and purpose of the survey and about a third (35 per cent) believed that it did not. Some of the questions asked by the latter group were: Who would know the results? How were the results to be released? Who would use the results and for what purposes? Most of this group, 30 per cent of all teachers, said that they had not seen any information from the ACER. This response relates to the manner in which the survey was handled in many schools. Three-quarters of the teachers who responded (74 per cent) said they had no opportunity of contributing to the decision of the school to participate in the survey. This response is consonant with the information gathered by the pre-testing questionnaire on the pattern of decision-making at the school level. In most instances the decision to take part in the testing was made by the school Principal and most class teachers were not involved in the administration of the tests nor in formal discussion about it. Consultation with staff was more common in government secondary schools. As observed earlier in the report, this appears to have been a deliberate strategy by some Principals to keep the survey "low key" and to avoid dispute. However, there was dispute in a few schools and in these schools staff opinion was strongly divided. Several schools withdrew from the testing program after initially agreeing to participate.

About three-quarters of the teachers who responded in the case study schools (73 per cent) believed that the school should be participating in the survey. Some of their reasons were: "information on national standards is useful for comparison"; "it can provide data for educational planning", "there is a need to assess the effectiveness of teaching". The minority opinion was that the testing served no worthwhile purpose as the reasons for the levels of performance were not examined by the survey and were, in any case, known to most teachers. Some suspected ulterior political motives for the testing and doubted that any action would follow that would benefit the students or teachers. There was frequent mention of the possible misuse or misinterpretation of the results taken out of context.

Very few teachers had any prior knowledge of the nature of the tests or the kinds of test questions that were to be given, although some were familiar with other tests of literacy and numeracy and assumed that they would be similar. It is clear from follow-up interviews with class teachers that no special steps were taken to prepare students for the tests. Teachers reported that most students were were unconcerned, although some were apprehensive about being tested.

One teacher said that he was now more aware of curriculum areas that needed attention, but almost all (99 per cent) of the teachers said there had been no effects on policy, curriculum, teaching or evaluation procedures in the school as a result of the project. In inteviews, most teachers said that the results would be unlikely to have any direct effect on their teaching (although some had got new ideas on testing techniques) but, depending on the use made of the results, there may be some later, and indirect, effects. Most teachers had little recollection of the 1975 literacy/numeracy survey and few could remember any school reaction to it. One teacher had at that time been on the staff of a school in a poor area of a regional city and the school had used the poor results in support of a claim for support from the Schools Commission as a "priority school". The same teacher said that the reaction had been to "tighten up the whole system" and to emphasize the "basics". By and large, teachers believed that their school was doing well and, in those schools where problems were recognized, there was little confidence that the survey would assist them in any practical way.

PARENTS

In the case study schools, the students involved in the ASSP testing program were also given a questionnaire to be taken home to their parents and, on completion, to be returned to the school. There was a very high response rate and a substantial measure of agreement in response to most questions.

Twenty-seven parents out of the 650 who responded to the questionnaire claimed that they were unaware that their child had been tested in the ASSP program. Almost all parents learned about the nature and purposes of the test from the printed material sent out by

the school, a few had attended school meetings where it had been discussed, and a number heard of the survey through their children, through radio talk-back programs, or media publicity about the opposition by teachers' unions. Although most parents were satisfied with the explanation of the nature and purposes of the survey, 20 per cent were not. They would have liked answers to questions such as: What is the precise purpose of the survey? Who would use the results and how would they be used? Was it evaluating children, the teachers, or the system? How would it benefit the child? Which schools were chosen? How were the children selected?

Ninety-one per cent of parents said that they had not contributed to the decision of the school to participate. Those who said that they had contributed said that they had done so by giving or withholding permission for their child to participate. Eighty-seven per cent of parents believed that the school should be participating in the survey in that it may be of benefit in detecting weaknesses, improving teaching methods, in providing a check on standards and assisting in educational research. Some saw some benefit in comparing the standards between government and non-government schools and between states, and others hoped it might result in a greater uniformity of standards across all schools. The 5 per cent who disapproved of participation felt that the survey would serve no useful purpose. There were more abstentions from this question than for any other, 8 per cent of parents writing in the words "no opinion" or "undecided". Perhaps they felt they were in the crossfire of opposing camps and wished to signify their neutrality.

Seventy per cent of the children were reported as having commented about the tests at home. Among the comments were:

"the tests were like simple IQ tests"
"they were easy"
"they were so easy they were silly"
"the newspaper items were fun"
"I was pleased to be chosen"
"why was I chosen?"

Others commented on particular test items that interested them. Other than the organizational details associated with the survey, no parent was aware of any special preparation for the survey carried out either by the school or their child.

STUDENTS

In those schools selected for case studies, interviews were held with groups of students who had been tested in the ASSP survey. There appear to have been few discernible differences in the responses from different types of schools or between urban and rural schools and (apart from a range of replies) between secondary and primary schools.

When asked why they thought they were asked to do the tests, a few students interpreted the question as meaning "Why were they, in particular, selected?" to which they responded "By chance" or "as a

result of random selection". More representative answers were:

"to test the ability of kids all over Australia"
"to find out what we've been taught"
"to find the standard of Australian school kids"
"to find out how bright we are"
"to see how they (the teachers) are teaching"
"to see if we know enough"
"to help the Government"
"to see if standards are falling"
"for research in education"
"to see if some kids are not coping with English and Maths"
"to help find the average across Australia"
"to see which school and state is the best".

Most students were told about the tests from two days to a week before they were administered, although in four schools they were told only just before the tests were administered. Letters to be taken home to parents, seeking their permission, were commonly distributed in class to the selected students by the Principal, Deputy or class teacher. In other schools, the students were called out of class to a special assembly where they were given some information about the survey and where the letters were distributed.

Most students were told that the testing would be in areas of English and arithmetic, that the tests would be paced, and that the testing would take about two hours. Many, particularly primary school students, were told not to worry about the tests; they would not be hard and would not count towards their school assessment. In one school the students said that they had no idea what to expect: "the test was just given out".

Having done the tests most students agreed that they were very easy. Very few said that they had any difficulty, but perhaps those who found them more difficult would be unlikely to admit to it in a group interview. Many said that they were interesting tests and commented on the novel items and, for them, the novel format: some said they were not used to booklets, cartoons and newspapers in tests. Some of the more sophisticated students found it hard to take the tests seriously. They found them to be ridiculously easy and the wait-time infuriating: one said "we've grown out of this". There was no evidence of any special preparation of students for the testing program.

Students were fairly evenly divided in their feelings of anticipation towards the tests: some admitted to being a little nervous or anxious, others a little excited, but a large number said they were not worried and had not given them a thought. A few teachers asked the students how they had fared, but most students said that their teachers did not say anything about the testing and some were not even sure that their teachers knew about it. In most schools the class teachers were not involved in the testing program in any way and it is clear that most were supportive or neutral, although a few expressed irritation at the interruption to their teaching and in a

few schools there was active opposition to the testing program by teaching staff.

CONCLUSION

Although opposition to the ASSP survey within the teaching profession resulted in some difficulties in sampling and selecting schools for case studies, a number of schools willingly participated. In these schools, which may not be typical, the testing program was carried out efficiently, with a minimum of fuss and with the general support of the parents. The decision of the school to participate was usually made by the Principal and classroom teachers were usually not involved in the testing program.

In the schools selected for case study most teachers believed that the school should participate in the study. Although many were dubious about the benefits and were concerned about the possible political misuse of data, few were explicitly opposed to the testing.

CHAPTER SIX

IMPACT OF ASSP ON PARTICIPATING SCHOOLS

Mal Hewitson and Colin Power

INTRODUCTION

In order to assess the initial impact which the ASSP test results had on participating schools, the evaluation team adopted two approaches. First, two structured interview schedules were developed for use in the case study schools, one schedule being for discussion with principals and the other with staff members nominated by principals as persons told to take some action following analysis of the test results. Second, a survey questionnaire was also developed for completion by other school principals who had originally agreed to take part in the evaluation of the ASSP project.

This chapter of the report, therefore, first presents an interpretation of the data provided by the principals and other staff members of the case study schools, and then reports on the responses to the survey.

The research instruments first asked respondents to report on the mechanical aspects of requesting and receiving their students' test results. "Initial impact" was then assessed in the following terms:

1. The emergence of significant factors in the analysis of student performance in reading, writing and numeration, and consequent implications for teachers and the school curriculum;

2. The effect on the school community as demonstrated by perceived pressure to divulge test results;

3. The outcomes of consultations, if any, by the principal with senior administrators or school staff;

4. Action actually taken as consequence of the test results;

5. Changes in lesson time allocation and teaching strategies in reading, writing, numeration or other subjects;

6. Actual effects on remediation activities within the school;

7. Stated firm intentions to initiate change directly attributable to student's participation in the ASSP project or their scores on the

tests; and

8. Other changes attributable to the school's participation in the project.

The findings of this phase of the evaluation are therefore grouped under these headings.

An early decision was taken not to approach the schools before late February, 1981. This time-lag was thought to be enough to allow test results to be analysed and make their initial impact on the schools. It was also thought desirable to allow the schools a few weeks to settle into the 1981 school year in order to gain their full co-operation. The data reported in this chapter, therefore, were gathered in late February and early March, 1981. Before reporting them, however, it is necessary to specify the sample of co-operating case study schools.

SAMPLE OF SCHOOLS

The case study schools agreeing to participate in this phase of the evaluation were distributed as follows:

TABLE 1: Distribution of Co-operating Schools by state/Territory

NSW/ACT	VIC.	QLD.	S.A.	W.A.	TAS.	N.T.	TOTAL
6	7	8	6	-	7	2	36

Table 2 shows the number of case study schools in the categories primary/secondary, government/non-government and urban/rural.

TABLE 2: DISTRIBUTION OF CO-OPERATING SCHOOLS BY LEVEL, CONTROLLING AUTHORITY AND LOCATION

Government Schools			Non-Government Schools			Grand Total		
Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
17	6	23	8	5	13	25	11	36
Primary:		12	Primary:		6	Primary:		18
Secondary:		11	Secondary:		7	Secondary:		18

Tables 1 and 2, however, include one primary and one secondary school unable to complete the interview schedule. In the former case, the two staff members given responsibility for the testing project had left at the end of the year. They had not retained a list of participating children nor actively sought printouts of results, and ASSP became no-one's responsibility.

In the latter case, a change of the principalship had occurred and no file on the ASSP project could be located by the incoming principal. It was thought that the previous principal, who had taken sole charge of the ASSP testing, may have filed the records among his personal papers.

In other schools where changes of staff had occurred, no similar problems appear to have been encountered.

Nevertheless, it cannot be claimed that the thirty-four respondent schools comprise a representative sample of schools by any of the criteria used to classify them in Tables 1 and 2. On the other hand, analysis of the responses of thirty-six case studies is certainly worth pursuing to determine what similarities and differences exist among them. For example, this could constitute prima facie evidence on which to draw some tentitative conclusions. Furthermore, data were received from an additional sixty-three school principals who completed the survey questionnaire. If these schools also reported the same sorts of experiences on the "impact" measures used, the case for generalising the findings to other schools participating in the project would be considerably strengthened.

With these considerations in mind, we now turn to an analysis of the findings.

FINDINGS - THE INTERVIEWS

In this section, the results of the analyses of the school principals' responses to the interview schedule are presented, and, where applicable, responses of other staff members. An assessment of the initial impact which the ASSP test results had on the schools is also included.

THE MECHANICS OF REQUESTING AND RECEIVING RESULTS

The first few items of the interview schedule asked whether the school had requested and received a print-out of student results and whether any delays or concerns had arisen in these respects. All but one school reported that results had been requested, but as the exceptional school had received the results anyway, it is possible that a request had been made and that fact overlooked or forgotten. As might be expected in view of the longer time required to score writing tests, the reading and numeration results reached schools relatively promptly, in many cases before the end of October, 1980. In the case of writing, a few schools reported receiving their results early in February, 1981, though one respondent interviewed early in

March claimed that reading and numeration results had not yet been received, and another that reading results were yet to come. Thus thirty-two of the thirty-four schools had received all results at least three weeks prior to interview, though one set of results had arrived only two weeks or so earlier in the case of two schools. Generally speaking, however, it is felt that the results had been with the schools long enough for them to have made their initial impact in terms of the measures incorporated in the interview schedule and survey questionnaire.

SCHOOL COMMUNITY PRESSURE TO DIVULGE TEST RESULTS

All respondent principals confirmed that test results had not been publicly divulged and strongly supported the principle of confidentiality. The only pressure reported was by one principal who had experienced what he termed "minimal pressure" on the part of one parent to disclose information on student performances. A number of respondents, however, disclosed their individual results to participating students, and one or two to interested parents.

On this measure of impact, therefore, it would appear that, at least initially, the ASSP test results had virtually no effect.

ANALYSIS OF TEST RESULTS AND IMPLICATIONS FOR TEACHERS AND THE SCHOOL PROGRAM

Twenty-nine respondent principals had read through their students' results and the other five had glanced at them in a preliminary way. Few reported any difficulty in understanding the print-out or the way results were set out, two finding it difficult as non-mathematicians and another reporting difficulty with the lay-out of results in the numeration sub-tests section.

An additional point was raised, however. One respondent could not see the purpose of the instruction to ignore the total writing score. There is no reason why this could not be briefly explained on the print-out itself in any future project.

Thoughtful analysis of the results was reported by twenty-one respondents, but among the thirteen who said that they had not really examined the significance of the results for the school and teaching staff, what tended to emerge (when pressed a little further) were responses very similar to those of the group of twenty-one. Thus:

(a) Almost all respondents remarked on the low difficulty level of the tests and the high scores of their students. The implication here was that no action on the part of the school or directive on the part of the principal was warranted on the basis of overly favourable ASSP test results. More importantly, however, it appears that many of the respondents were thinking in terms of scores on achievement tests which differentiate among and within good, satisfactory and poor results, and not in terms of basic competence mastery tests which differentiate only between masters

and non-masters.

(b) Along with comment on the high scores of students, most respondents stated that the ASSP test results confirmed what was already known about student abilities in the areas tested, the implication being that no new significant factors had emerged to justify any change in the school curriculum or teaching strategies.

(c) Many respondents pointed out that some yardstick was needed before the ASSP results could be meaningfully interpreted. One third said that they were awaiting the publication of state and national results with which to compare their students' performance on the tests, and others observed that, without knowledge of the cut-off scores below which students would be judged "not minimally competent", the results were not useful as an input to school policy decisions on reading, writing and numeration. Information on cut-off scores could readily be incorporated in the print-out of results in any future project.

(d) One quarter of the respondents indicated that, irrespective of the results, they would not be prepared to initiate significant change because of the smallness of the sample of students who participated in the ASSP tests. The implication here may be that the possibility of item by item analysis in order to diagnose areas worthy of further investigation was overlooked by some respondents. On the other hand, the existence of remedial programs in a number of the schools tends to suggest that diagnostic activities were already appropriately catered for.

(e) It was pointed out by at least one respondent that the need to obtain parental permission for the student to participate in the ASSP tests could bias the sample. For example, it was stated that the parents of some of the lower ability students refused permission, and more able students were selected in their place.

In light of the above factors, it is not surprising to find that respondents agreed with and were satisfied with the results which their students had obtained on the ASSP tests. Their main reservation had to do with the low difficulty level of the items and, for the great majority, the print-out results were adjudged to be of little practical significance or value. Thus on this measure of impact, the initial effect of the ASSP test results was negligible.

CONSULTATIONS WITH OTHER STAFF AND THEIR OUTCOMES

While not one of the thirty-four respondents had discussed the school's ASSP results with superordinate administrators, twenty-four of them stated that they had consulted other school staff members about the results. Teachers of participating students were consulted in fifteen schools, the Deputy Principal or/and Year Co-ordinator in eight schools, other senior staff in nine schools, and other teachers (including discussion in the general staff meeting in two instances)

in six schools.

In most schools, nothing definite emerged from these discussions beyond "having another look at a later date". For example, there were cases when the Year Co-ordinator or equivalent had been asked to check through the results to see if anything of value emerged. Again, the results were seen by two respondents as giving impetus to an already developing emphasis on reading and comprehension; similarly with regard to a remedial program in another school, and an arithmetic program in yet another. Two respondents felt that the results would form useful additional input to a general review of the curriculum planned for later in the year.

ACTION ACTUALLY TAKEN

In developing the interview schedule, the evaluation team felt that it would be useful to know specifically who had been told to do what following analysis of the test results, and then to check with the individuals concerned to determine whether the proposed action had or had not been taken.

In the event, however, there is little further to report on this measure, beyond what has been said above. Those persons nominated by principals as staff members told to take some action (such as Deputy Principals, Year Co-ordinators, etc.) were approached and in all cases confirmed what the principal had said. Those who had completed their task of reviewing the students' results and reporting back to the principal had recommended that no further action be taken.

Findings on this measure, therefore, support the conclusion that, apart from isolated instances, the initial impact of the ASSP results was, at most, slight.

The fact is, then, that at a point in time some four months after the administration of the tests, only two instances were reported of change having been actually implemented in any of the thirty-four respondent schools. One change was relatively minor - the principal of a primary school instructed the teachers concerned to give more emphasis to a particular segment of numeration work involving digital and clock face time.

The other change, however, was major. Having received the test results promptly and analysed and discussed them with teachers of participating students, one primary school decided to increase the lesson time allocation to reading, writing and numeration from the beginning of Term 1, 1981, at the same time reducing (with some regret) the time allocated to other subjects. Accompanying this change was the adoption of a more formal, traditional approach to teaching.

One interesting spin-off from the test results was the adoption by one teacher of the ACER format for a course in computer usage. Generally speaking, however, it seems fair to conclude that on this measure of impact, the initial effect of the ASSP test results was, at

most, slight, with one isolated exception.

CHANGES IN LESSON TIME ALLOCATION AND TEACHING STRATEGIES

Among the thirty-four respondents, changes in terms of lesson time allocation and teaching strategies had occurred as a result of the ASSP test results in one primary school. While the results had made a significant impact (according to this measure) in the one school, their impact otherwise was nil.

ACTUAL EFFECTS ON REMEDIATION ACTIVITIES

No demonstrable effects on remediation activities were reported by any respondent. One lacked resources to initiate a remedial program which the ASSP test result showed to be desirable, and in other cases, added impetus was given to existing remediation programs.

STATED FIRM INTENTIONS TO INITIATE CHANGE

Although little change may have been actually implemented by respondent principals, it was still possible for them to be planning change which could be directly attributed to students' participation in the ASSP project or their scores on the tests. Principals were therefore asked whether they had firm intentions to initiate such change in the near future.

Five of the thirty-four principals stated that they had firm intentions of introducing some change as a result of the ASSP testing program. The envisaged changes included more emphasis on the teaching of reading and comprehension and on "thinking type" questions, more attention to reading and writing, and more use in two schools of newspapers as a teaching resource. On this measure, too, it would thus appear that the initial impact of the ASSP test results on the thirty-four schools was slight.

OTHER CHANGES ATTRIBUTABLE TO PARTICIPATION IN THE ASSP PROJECT

The final item on the interview schedule sought any other comments which respondents might care to make on the ASSP project. As far as initial impact of results is concerned, the most frequent remark was that those teachers who had seen the results were pleased with their students' performances. The same was true of parents.

Clearly, however, participation in the project did not always serve to promote staff morale. It seems that some principals consciously kept the project as low key as possible. Because of divided staff opinion on the issue of school participation, it was felt that the less impact the test results had, the better. Perhaps the principal who had transferred to another school and filed the ASSP records among his personal papers was reacting to this sort of pressure from staff or from union representatives. In another school,

the principal stated bluntly:

> There was almost total staff objection to the tests being given which...caused a lot of trouble, loss of morale in the school. For this reason the results have been kept in very low profile.

Given the abrasive context in which the ASSP project was at times implemented, it is difficult to assess the impact of the test results on such measures as staff morale since most respondents made no reference to it at all. But there was a strong negative impact in one of the case study schools. In this school, the principal and staff were in sharp conflict over ASSP.

Although not to do with impact measures, a number of comments of a general nature made by the respondents appear to be worth recording briefly. It was remarked, for example, that the ASSP project was a "political exercise", using the term in its perjorative sense. In another case, the project was perceived as a means of bringing pressure to bear on teachers, and in another, as the beginnings of a system of payment by results. Others regarded the ASSP tests as superfluous, first, because teachers already knew what the results told them; and, second, because the ASSP tests duplicated either school or State-wide tests which encompassed not a small sample but all students, thus making results much more useful for school purposes.

The inevitable time-lag between taking the tests and receiving the results was another factor mentioned which may explain in part the negligible impact which the ASSP test results had on the schools. When results were received, they seemed "detached" from the ongoing classroom activities, and even more so if changes to the composition of students or staff had intervened in the meantime, e.g., over the Christmas holidays.

A few respondents also stated that they agreed to participate only in the interests of helping research, and irrespective of whether projects such as ASSP were mounted or not, their staffs were endeavouring to improve their teaching effectiveness all the time.

Among those who were awaiting publication of national and state results as criteria for comparative purposes was one respondent who anticipated that the forthcoming information would be useful for inter-school comparisons.

Finally, one respondent criticised the tests on the grounds that the sample of skills selected for testing were not always in line with school objectives and that some of the items were out of date.

All the issues which these comments raise are considered elsewhere in this report (see Chapters Three and Four). But it is pertinent to note here that as some respondents' expectations of the ASSP tests were substantially different from the stated purposes of the project, the vast area available for misunderstanding may partly explain why the initial impact of the test results was minimal, save for isolated cases.

INTERVIEWS WITH SCHOOL STAFF MEMBERS

In the thirty-four case study schools, relatively few staff members were nominated by principals as persons told to take some action following analysis of the test results. As reported above, only two instances were reported of change having been actually implemented by late February, 1981. Staff members interviewed invariably confirmed the responses which principals had already given, and no new information was elicited during this part of the initial impact evaluation.

FINDINGS - THE SURVEY

The findings with regard to the case study schools are supplemented in this section by the responses of principals from other schools agreeing to participate in the evaluation of the ASSP project. The survey was conducted in late February, 1981. The 63 responses received represented a 37 per cent response rate. The results, therefore, should not be assumed to be representative of all participating schools.

The outstanding feature of the survey responses is that they follow almost identically the patterns established by the interviews conducted in the case study schools. For this reason, this segment of the evaluation is reported here only in summary form.

Nearly all schools had requested and received their ASSP results with no problems arising, but five respondents indicated they had not yet studied the print-outs, even though they had been available for some time. Some had difficulty in interpreting the results in that the criteria for scoring were not included on the print-out.

With regard to school community pressure to divulge results, three respondents reported that they had been subjected to "a little pressure" to release the school's results and that they had strongly resisted such pressure.

Analysis of test results resulted in no change as far as teachers and the school curriculum were concerned. Both the small sample size and the fact that selected students came from a number of classes were suggested as reasons for the lack of impact of the ASSP test results. In addition, in most cases, the results were seen to contain little new information and were either as expected or higher that expected. In only one secondary school were the results lower than expected.

Those principals who had consulted with others about their students' results - about half of the respondents surveyed - had mainly held discussions with the deputy, other senior staff, and the teachers concerned. However, in three schools, results had been discussed with the School Council or the Parents' and Citizens' Association.

Little action was initiated in any of the respondent schools. In one, teachers had been encouraged to work towards developing library skills as a result of the tests, in another school, the test item on digital and clock face time indicated a weakness to be rectified, and in a few others, some increased remedial activity in reading and writing had resulted.

At the end of February, 1981, therefore, it was evident that the initial impact of the ASSP test results had been minimal among the schools surveyed. Furthermore, there was little evidence among the respondents of firm intentions to make changes in the near future as a consequence of the ASSP tests. The most frequent commitment of respondents in this respect was that of awaiting publication of the national and state results later in the year, at which time their students' results would be more meaningful and, depending on their comparative performance, could have an impact.

It is important to note that if there is a bias in the survey sample, it is likely to be in the direction of principals sympathetic to national assessment. It could thus be argued that the initial impact of ASSP test results in schools whose principals were not sympathetic to national assessment would probably be weaker rather than stronger. On all counts, therefore, it seems fair to conclude that the ASSP test results had minimal impact on participating schools except in very isolated cases.

CONCLUSION

It will be recalled that in Chapter 3, it was argued that testing a light sample of students once every four or five years in specified areas of basic literacy and numeracy would be unlikely to have a marked influence on the school curriculum and teaching methods. In terms of the initial impact of the ASSP test results, as measured in the present evaluation, such a conclusion is substantiated. In short, in both the case study schools and other schools participating in the evaluation, the impact was mainly negligible and, at most, slight, save for one or two isolated instances.

On the other hand, this phase of the evaluation of the ASSP project has brought to light a number of additional points of importance.

First, teacher union support appears to be a prerequisite if similar ventures in the future are to be fully successful.

Second, efforts need to be made to eliminate the widespread misunderstanding of the nature and purposes of such tests, revealed by statements of surprise or disparagement at the high test scores of "relatively weaker" students.

Third, the usefulness of item by item analysis of results needs more explanation and greater emphasis.

Fourth, the derivation of minimum competence scores and the cut-off score in each of the content areas tested should be incorporated in the print-outs of results.

Fifth, the time-lag between testing and receipt of scores by schools should be reduced. Alternatively, the tests could be administered earlier in the school year to reduce the effect of changes of teachers and class members during the time it takes to score the tests and return the results to the schools.

Finally, it needs to be reported that, despite the hostility of some groups towards the whole project, the interviewing teams found considerable goodwill among school principals who participated, and the survey revealed that just over 50 per cent of respondents considered that a continuation of national testing would be beneficial. Seventy-five per cent gave support or qualified support for testing at the national level by ACER and the Australian Education Council. Only four respondents said "Never again". There was very strong support for programs like ACAP designed to facilitate assessment of basic skills at the school and classroom level.

CHAPTER SEVEN

REACTIONS TO THE RESULTS OF TESTING

Donald Hogben, Malcolm Eley and Colin Power

INTRODUCTION

The results of the 1980 national testing became available to the AEC and its advisers in the middle of 1981. On the 9th September 1981, these results were made public in a press release from Mr. W.A.M. Gunn, Queensland Minister of Education, the then Chairman of the AEC. The press release was accompanied by a five page statement containing extracts from the report of the national results (1). This report, written by ACER staff was made available through the Australian Government Publishing Service. In addition, the AEC distributed a short brochure describing the project and summarizing the results.

In this chapter, we report reactions to the national results by politicians, the media, education departments, unions, principals, teachers, parent and community groups.

THE 1980 NATIONAL RESULTS

The ACER Report provided estimates of the percentage of 10 and 14 year-old students throughout Australia who, by the criteria employed by the ACER, could be classified as being competent in each of the basic skill areas tested. For example, the report indicates that (i) performances of 10 and 14 year-olds were 70 and 72 per cent respectively for comprehending a prose passage, 99 and 100 per cent respectively for comprehending a simple sentence, and 71 and 83 per cent respectively for the total reading test; (ii) writing performances ranged between 57 and 93 percent for 10 year-olds and 54 and 97 per cent for 14 year-olds; (iii) the numeration test was "mastered" by 78 per cent of 10 year-olds and 86 per cent of 14 year-olds. The report also shows that girls out performed boys on most tasks, the only area assessed in which boys were superior being the reading of maps, graphs and scales at the 14 year-old level.

The ACER found it neither "necessary nor desirable" to make judgments as to whether the performances reported were satisfactory or not, although the report does seem to suggest that the percentage of 10 year-olds correctly answering items concerned with time and computation ought to be higher. The one aspect of the results which

was regarded as particularly noteworthy was the comparison between the 1975 and 1980 results. The authors conclude:

> There was no significant area in any of the assessments made where student performance at either age level has declined in the five years between 1975 and 1980. In fact there were a number of specific areas where student performance was higher in 1980 than it had been in 1975. In particular, the better results in 1980 for newspaper reading by 10-year-olds, for comprehending prose by 14-year-olds, for writing a personal message by 10-year-olds, for the basic use of the four arithmetic operations by-10 year-olds, and for a range of other numeration tasks by 14-year-olds should be noted (2).

POLITICAL REACTIONS

The press release issued by the AEC reveals that the Ministers of Education were clearly pleased by the results. Speaking on behalf of the AEC, the Chairman (Mr. Gunn) announced:

> What is particularly gratifying to me and my Ministerial colleagues is to read that not only has there been no decline in standards of performance in any significant area over five years, but that there have been improvements at a number of specific points (3).

The AEC releases reaffirm the importance given by the Ministers to the skills tested and the value to the Ministers of data which "provide a significant indication of the state of basic skills in our schools". From their perspective, the results could be seen to dampen the standards debate, and to provide evidence that national and state educational efforts aimed at improving basic skills were working. In addition, the results were seen as providing a useful basis for directing attention to areas in need of further study; for example, research on the reasons for the different performance of girls and boys in the areas tested.

Whereas the stress in 1975 was on the weaknesses in performance, the ministerial statement in 1981 focused on improvements in basic skills: improvements in the understanding of newspaper articles, writing a personal message and the basic use of the four arithemetic operations by 10 year-olds, and improvements in comprehension of a prose passage and numeration tasks (including reading maps and scales, calculations and using percentages) by 14 year-olds. The AEC summaries noted that, at the 10 year-old level, girls generally out performed boys. They also highlighted the conclusion from the report that, measured against the prose-comprehension performance of 14 year-olds in other English speaking countries in 1970, 14 year-olds in Australia now performed at a higher level than students in England and equal to those in the United States and New Zealand.

The flavour of the AEC releases was positive. However, some aspects of the performance of 10 year-olds were seen to be in need of attention by primary schools·

> Only 72 per cent of 10 year-olds could tell the time and work with time and only 67 per cent of 10 year-olds could compute effectively using the four arithmetical operations (4).

The ministers considered it essential that all students in normal schooling should be able to perform these tasks before leaving primary school. While most politicians, educators, parents and employers would agree, what would represent an acceptable proportion (75, 80, 85 or 90?) for a given task and age group is unclear. Standards are not absolute but are relative to the framework of values, aspirations and understanding in what is involved in achieving and measuring them of various groups within society at a given point in time.

In the national parliament, the results of the 1975 survey were cited on a number of occasions in discussions on standards of literacy and numeracy. In contrast, the results of the 1980 survey have been virtually ignored. By 1981 Commonwealth government policy towards, and funding of, tertiary education and government (vs private) schooling displaced standards as a political issue. The cuts announced by the Prime Minister on 30 April (following the Report of the Committee of Review of Commonwealth Functions), and subsequent statements and bills introduced by the Commonwealth Minister of Education (Mr. W. Fife), were strongly criticised by the opposition, brought loud protests from many educators, but were generally supported by employer groups and the major daily newspapers. It was in this context that, during the second reading of the State Grants Bill, passing reference to the results of the 1980 survey was made by the opposition spokesperson for education, Mr. J. Dawkins:

> What the Prime Minister overlooked is that the innumeracy and illiteracy of school leavers is not a new phenomenon. It has been with us for as long as we have had school leavers. Of course it is unacceptable and of course something has to be done about it, but he also overlooked that the position has been greatly improving. The levels of literacy and numeracy of school leavers and, indeed, of kids still at school have been gradually improving over a long period. The most recent evidence of this comes from the report of the Australian Council for Educational Research...that study reveals that there has been a very significant improvement in relation to these problems of illiteracy and innumeracy amongst kids at Australian secondary schools (5).

At the 1982 Annual Conference of the Australian Teachers Federation, Mr. Dawkins again used the results of the 1980 testing in an attack on the Commonwealth government and the Prime Minister. He accused them of damaging public confidence in schooling, ignoring the fact that the problems relating to the lack of basic skills amongst some school leavers are of long standing, and of failing to acknowledge the evidence from the ASSP project.

The feared recurrence of the political attack on schools following the release of the 1980 results was not forthcoming. The results were used by the Ministers of Education to defuse the standards debate and to vindicate their position, and by the Labour party in its attack on the Fraser government's educational policies. But compared with the 1975 results, the political impact of the 1980 results has been slight indeed.

MEDIA REACTION

Reference was made in Chapter two to the extensive coverage given to the results of the 1975 testing program. Led by the Bulletin (to whom a draft of the ACER Report was "leaked" in 1976), primary and secondary schools across Australia were attacked by the media. The reporting was frequently sensationalist in style, and many claims were made about the failing education system and declining standards.

It has been a common observation that the media thrives on bad news and sensation, and the results of the 1975 Australia-wide testing of 10 year-old and 14 year-old children were certainly seen as news: the coverage was quite extensive. In its report on the study, the ACER pointed out that the public debate following the release of the results had contained much misinterpretation and indeed misrepresentation of the actual results of the study. Claims were made that standards of literacy and numeracy were in decline but, in the absence of any comparable data from a previous time, these claims were difficult to sustain. As has been noted earlier in the present report, the 1980 ASSP project sought, among other things, to provide data which would enable direct comparisons to be made with the results of the 1975 testing.

In view of the prominence given by the media to the 1975 test results, and acknowledging the influential role played by the media in public debate about education, the evaluation team considered that a monitoring of the media response to the 1980 testing was an essential part of the ASSP evaluation. Consequently, the press, radio, and television were closely monitored by the evaluation team from the date of the press release from the Australian Educaion Council.

In marked contrast to the media coverage given to the results of the 1975 literacy and numeracy testing, the September 1981 results release produced a very limited media response. It might not be stretching things too far to refer to the appearance of the results as a "non-event".

The results were virtually ignored by the television networks. On the day of the release of results, the Secretary of the AEC (Dr. R. Browne) was interviewed by ABC News in Melbourne. One hour after Dr. Browne had stated that the results showed no declines and improvements in some areas, the ABC in "This Day Tonight" contained a segment on the provision of resource teachers in Victoria "to deal with the massive decline in literacy and numeracy". There was a short interview session during the "Nationwide" program on ABC TV in South Australia on September 10. The interviewees were Dr. Peter Tillett,

a Deputy Director-General in the State Education Department, and Geoffrey Partington, a Senior Lecturer in Education at Flinders University. The program was reasonably balanced and fairly low-key. Dr. Tillett saw the results in a positive light, but did make the point that the schools are stressing much more than the "basics". Mr. Partington, while agreeing that there had been some improvements in performance since the 1975 testing, suggested that with improved teacher pupil ratios, and better school conditions generally, even higher levels of performance might have been expected. He criticised the survey for the absence of inter-state comparisons, and for the lack of group comparisons apart from comparisons by sex.

ABC radio carried straightforward accounts of the survey's major findings during its news broadcasts on the evening of the release of the report, and again the following day. No detailed consideration of the report was broadcast until "Education Now" on the first of October. During a segment occupying some 13 minutes, interviews were conducted with Professor Hedley Beare, a Professor of Education at the University of Melbourne, and Graham Little, a Principal Lecturer from the Canberra College of Advanced Education. Professor Beare referred to the way in which the 1975 results were used by a number of people to attack the schools, and he expressed the hope that the results from the 1980 testing would help to restore confidence in the schools, which, he suggested, were doing an "honest job". Nevertheless, he urged caution in interpreting the results, noting that the test items dealt with only a part of the curriculum of schools. In his words: "We ought not to be too satisfied about using literacy and numeracy as our key indicators of children's progress ...They deal only with a small part of the school's curriculum".

Graham Little was more critical of the survey itself, and the employment of externally administered tests in the determination of the educational achievements and specific needs of particular children. He saw more benefit attaching to internal assessment by competent teachers. As he put it:

> I think that the ACER survey as such probably has some impact and relevance at the very broad level of national or perhaps political debate. And there's an end to it. But they [such surveys] don't tell me as a specific teacher in a specific place anything.

Press coverage of the ASSP report was considerably wider than that given by radio and television. Despite this greater attention, the coverage was anything but extensive and the variation from state to state was considerable. The West Australian carried a summary of the AEC press release in its September 10 edition. In South Australia, coverage was limited to one short article on page 10 of The Advertiser on September 10. The Advertiser article was a piece of factual reporting of the major findings with no inferences made nor interpretations offered. The article concluded with a comment from the Director-General of the S.A. Education Department, who was "very pleased and gratified" by the results. The article referred to the fact that the Minister of Education had noted that no comparisons would be made between schools or states. Reference was also made to

the independent evaluation being undertaken.

The Brisbane Courier Mail carried a short, factual report headed "Girls read better - survey" on page 14 of its September 10 edition. The Courier Mail's report contained no interpretive comments and it placed some emphasis on the generally significantly higher attainment of girls on the reading and writing tests. No further material appeared in the Courier Mail apart from a letter commenting on sex differences which was published on September 15 (The same letter appeared in The Australian on September 16).

The Brisbane afternoon paper The Telegraph published an educational editorial on September 10, not reporting on the ASSP results, but rather offering comment on allegations made before the release of ASSP results by Professor Chipman, Professor of Philosophy at the University of Wollongong, about the low teacher and school standards in N.S.W.

In N.S.W., the Sydney Morning Herald (4.9.81) and the Daily Telegraph (4.9.81) devoted far more space to reporting on the allegations made by Professor Chipman than to the ASSP results. This is not to say that the Chipman allegations were uncritically received, however. The Sydney Morning Herald editorial of September 5 was clearly critical of Professor Chipman, accusing him of "headline hunting" and "rabble-rousing". In a letter published in the Sydney Morning Herald on September 14 Professor Chipman, in the course of defending his allegations of declining standards, commented that

> "...The evidence which should worry us all about the quality of public education is far more fundamental than the comparative reading scores of 10 year olds today and five years ago. It is evidence about where the thinkers and policy makers within education systems wish to take us, as shown in their publicly available writing".

Two letters appearing together with the Chipman letter employed the ASSP results to counter Professor Chipman's allegations. Clearly these letter writers drew quite different inferences from the survey findings.

The NSW Sun Herald published a two-part series on the ASSP testing, in which the paper presented details of many of the test items from the numeracy, reading and writing tests. In addition to the presentation of a selection of items, the series offered some comment. In the first article, it quoted the NSW Director-General of Education at some length in his support of the ACER's test designing capability and in his belief that the results demonstrated no decline in literacy and numeracy skills.

In Victoria, The Age published a short factual article on page three of its September 10 edition under the heading "Study awards pupils top marks in 3Rs". This short article was followed in the September 22 edition by an editorial and an article by Professor Hedley Beare titled "Literacy and numeracy: popular myth destroyed". The Age editorial noted that the survey revealed no decline in

performance in any significant area since 1975 and that there have been improvements in several areas. However, it suggested that parents and teachers must be concerned"...that substantial numbers of our young people do not have some of the abilities in literacy and numeracy most of us would regard as essential". The editorial then provides some examples including the finding that"...among 10 year-olds four out of 10 cannot correctly multiply a two-digit number by a one-digit number..." The editorial commended teachers for their work, but wondered about the maintenance of progress, particularly if expenditure on the schools is allowed to decline.

Professor Beare in his article in The Age drew attention to the controversy surrounding the conduct of the ASSP survey, and he suggested that this may account for the fact that the generally positive results prompted "no cheering from the grandstand". Professor Beare noted that there was an outcry around the country five years ago, and asks "Where are the schools' detractors now?" He believes that the massive expenditure on education has paid off in quite tangible ways, and wonders if parents will now be satisfied that the standard of achievement in the nation's public schools is steadily rising. He observes that the report itself makes it clear that the survey was politically motivated and conjectures on what politicians will make of the results. Will funding for education be stepped up, or will present levels of funding be maintained?

He also suggested that some of the improved results obtained in 1980 resulted from teachers having concentrated on certain weaknesses revealed by the 1975 testing. Professor Beare, however, raised some doubts about "mastery testing" of 10 and 14 year-olds, and the levels set, suggesting that attention needs to be paid to children who are incapable of achieving the levels required.

Approximately one month after the release of the ASSP results, The Bulletin published a full-page article headed "Quality of teaching the x factor" written by Professor Chipman (6). In the article Professor Chipman pointed out that, contrary to much public comment, the report does not show that there has been no general decline in average level of school achievement. He drew attention to the fact that the survey was concerned with basic levels of achievement, and notes:

> While it would be churlish not to welcome the report's evidence of the rise in level of minimum competence, it would be a blatant error to cite it as proving, or even claiming, anything else.

Professor Chipman devoted the rest of his article to an aspect of the survey and the report which he considers to be very significant; the absence of comparisons between students from government and non-government schools. Referring to the poorer performance by students from government schools in 1975, he pointed out that the 1980 results do not allow any conclusions about improvement or decline in state schools since that time.

Jim Coates, a teacher at Copland College, drew a quite different conclusion about kind of school attended in an article entitled "Are standards falling? Don't you believe it!" which appeared in the Canberra Times in late October (7). Mr. Coates outlined some of the survey's major findings, pointing to a number of performance increases over the 1975 results. He then observed: "These improvements could have only occurred if children in government and Catholic schools had done better in 1980".

Mr. Coates sees little cause for complacency, however. He pointed out that, despite the improvements since 1975, "...a significant minority of the 10 and 14 year-old children tested in 1980 did not master the most difficult items in the survey ...". He also drew attention, as did Professor Chipman, to the fact that the survey concentrated on "...the basic foundations of literacy and numeracy in 10 and 14 year-old children and did not measure higher levels of skills older students seeking employment would be reasonably expected to possess".

The Australian published a short, factual report of the ASSP results on the day following their release. The report summarized some of the major findings and quoted from the AEC press release (the words of Mr. Gunn quoted earlier in this chapter) and briefly from the ASSP report itself. The article offered no interpretation, but did conclude by observing that "despite the high success rate, only 72 per cent of 10 year-olds could tell and work with time, and only 67 per cent could effectively use arithmetic".

In a short editorial comment the following day, The Australian (11.9.81) pointed to the exceedingly complex nature of considerations of educational standards. While welcoming the 1980 results and the improvements over the 1975 findings, the paper cautioned against overestimating the importance of the findings, pointing to recent surveys which have shown "...significant, even critical, deficiencies in university applicants and in future and present teachers ...standards may have fallen for select social groups while overall standards have been maintained. We don't know". Having cited what it considers may be one of the most important gains since 1975 - mastery of newspaper items (probably reflecting increased emphasis by schools on newspaper reading) - the editorial concludes: "But we still have a long way to go to fit children for the sort of job market they will face in the 1980s".

Among the small number of letters published in The Australian, one from Dr. Eva Eden, President of the Australian College of Education, was mildly critical of the ASSP survey and similar basic skills testing programs. The letter pointed out that such surveys do not reveal why improvement or decline occurs, nor do they indicate how improvement can be fostered. The letter also pointed out that "...large-scale testing of minimal level skills of themselves contribute [sic] little to identifying and developing in children the range of skills needed to cope with a more complex and sophisticated world" (8).

EDUCATION DEPARTMENTS' REACTION

Approximately one month after the release of the ASSP results, reactions were sought from key education department officials in three states: Queensland, South Australia and Victoria. What follows is a summary of the telephone interviews held.

Queensland

The Education Department clearly appreciates the limited scope of the ASSP testing, and is also fully aware of the opposition from several quarters to the whole testing exercise.

The Department has no intention of making any public statements about the survey's findings. It does not wish to provoke any further confrontation over ASSP in Queensland. Copies of the Summary Report have been sent to schools and regional officers. Additionally, multiple copies of the full Report have been placed in central and regional departmental libraries for reference.

Through its specialist branches - especially Curriculum Section, Research Branch, and Information and Publication Branch - the ASSP findings and their implications for formal Syllabus Committees and associated bodies at both primary and secondary levels will be opened for discussion. Their relevance will depend on Syllabus Committee perceptions, and it is at this level of operation that the main impact of the results is likely to be realized. The Education Department regards the ASSP report as a curriculum document. Inspectors of schools will be expected to deliberate on the implications of the report for their advisory role in schools.

The Education Department proposes to take steps to have the Queensland results further analyzed to produce more detailed information. The intention is to compare the state results with the national findings and to pass on the results of these comparisons to relevant persons and bodies.

South Australia

Two key Education Department officials were separtely interviewed. Since there were no significant differences between them, what follows is a summary of their combined comments.

The Department does not propose to take any particular action as a result of the ASSP survey and, in fact, it has deliberately played the whole exercise "low key". In South Australia the project has been somewhat of a "non-event": there was little enthusiasm for it, nor active opposition to it.

The results are generally seen as reflecting the fact that the schools are performing well, and improvements over 1975 are welcomed.

At the present time the S.A. Education Department is discussing state testing and what emphasis will be given to school-based developments as contrasted with centrally-focused programs. Presently, South Australia is the only state in Australia which has no state-wide testing program of some kind. There is, however, a lot of testing activity in schools.

There are no plans at present for the state data to be analyzed by the Education Department, although it does hold the data tape. In its September edition, the S.A. Education Department staff newspaper, Inside Education, carried a summary of the findings presented in the ASSP report. The article was headed "National survey finds no decline in standards". Most of the article was devoted to straight, factual reporting of the main results. A final section presented the views of a S.A. Education Department Superintendent, of Curriculum, Mr. D. Pallant. Mr. Pallant served as a member of the ASSP national steering committee. His comments indicated a lack of enthusiasm for the project generally, and that he (and presumably the S.A. Education Department) favours testing programs much more in the hands of teachers. As he put it:

> As a general indicator, it [ASSP] is not too bad. It tells us not that we have reached a pinnacle, but that the claimed rapid decline is not true...The survey also suggests to me that there are probably other areas, such as the Australian Cooperative Assessment Project, in which we could invest our energies if we want to do something on a national level. While I have some personal reservations about that project, it is an opportunity for states and school systems to cooperate in the development of good assessment devices and procedures that would be available to schools and teachers to use in ways they see fit (9).

Victoria

The ASSP study and similar projects which might be mounted in the future, are not viewed with enthusiasm in Victoria. In fact, among the senior education department officers across Australia there is the view that ASSP should not be repeated. It appears that not only in Victoria (and in South Australia, as noted above), but generally across the states, there is far more support for the Australian Cooperative Assessment Project with its emphasis on diagnostic, progress and review testing.

The view was expressed by the Victorian interviewee that the distrust of the ASSP and similar surveys among teachers and administrators is not simply a reflex defence mechanism, but it is based on the widespread experience gained overseas. That experience is that large scale testing tends to result in constriction of the curriculum. Examples cited were the 11-plus examinations in England, and the NAEP and basic skills testing in the United States. However, the interviewee believes that the ASSP will be a "non-event" in terms of its influence on curriculum testing or policy in Victoria. It has, he considers, merely confirmed that most children are learning the basic skills and that we have problems with the bottom 20 per cent of the student population. These problems are to be expected given the

variability in any population. However, in his view, politicians and employers are not likely to admit that schools are doing well and that the problems are in fact not of their making.

Union reactions

The 1980 ASSP project and the release of the results were highlighted at the 1982 Annual Conference of the Australian Teachers Federation by its president (Mr. G. Tickell). In his opening speech he claimed that there are a number of direct political pressures at work to shape the curriculum of schools, one of the most significant being national assessment. He referred briefly to the union boycott of ASSP, claiming that this was a "considerable achievement" but not a complete success. The lack of support from some members was attributed to the "naive, but understandable view" that this was a professional, non-political issue. In the ATF view, the immediate political implications of ASSP are not clear. On the one hand, the ATF is concerned, as it was in 1980, about the risks of selective reporting and political uses of national testing. On the other hand, it wishes to be seen as willing to be held accountable:

> Certainly we have nothing to fear from properly conducted tests which can only reinforce the overwhelming evidence that levels of literacy and numeracy are improving (10).

Generally, the President of the ATF is confident that national testing will serve to "strengthen our hand in narrow political terms".

While the level of support for, ASSP by the ATF seems to fluctuate depending on the degree to which the results can be or have been used to defend or attack the teaching profession, the Victorian Teachers Union remains adamant in its opposition to ASSP. In response to our questions, the Union general secretary (Mr. B. Rimmer) indicated the reasons for its policy and its reactions to the report. His letter stated:

> Considering the amount spent on this project, a prime concern must be the question of educational benefit. The results reveal nothing about individual children, nothing about why such results were obtained, and most importantly, nothing about what educational strategies could be implemented for the educational advantage of Australian children (11).

His letter went on to raise questions about
(i) the validity of the results (given that the achieved samples were well below the designed samples) and of the ramifications of using a weighting statistic to correct for biases due to the withdrawal of government schools;
(ii) the validity of the tests constructed (given what the VTU sees to be inadequate clarification of the concepts of literacy and numeracy and the limited exploration of the nature of the reading process);
(iii) the validity of comparisons between "standards" in 1975 and 1980 (given that the number of items required for mastery on reading sub-tests differs for each age level; the criterion of "successful

performance" in 1975 was taken to be 80 per cent while the criterion for comprehension sub-tests for 10 year olds in 1980 was three out of five).
The VTU continues to have grave doubts about the results and believes that national testing programs like ASSP are of "very limited use" and an "unwarranted expenditure".

VALIDITY OF COMPARISONS

In Chapter Four, the technical issues relating to sampling test development and establishing national estimates were addressed. One must ask as well about the validity of the conclusions reached regarding differences in performance associated with background variables (sex, school location, year level) and year of testing (1975 and 1980).

The Tukey-Quenouille Jacknife procedure was used by the ACER to estimate the standard errors for items, comparisons between 1975 and 1980, and comparisons between sexes. Having examined the results of breakdowns of student performance by sex, location and year level undertaken, the evaluation team is satisfied that the conclusions reached by the authors of the ACER report regarding sex, location and year level differences were justified.

In comparing the performances of students in 1975 and 1980 on the 27 common or "link" items and the two common writing tasks employed, the same statistical procedure was used. Again, the evidence does support the major conclusions reached in the report: there was no significant area in which student performance was lower than in 1975, and there were a few specific areas where student performance was significantly better than in 1975. However, we would question one conclusion given some prominence in the report and the media , namely that there was a "marked improvement" in the performance of 10 year olds on the newspaper items. It seems to us that the degree of improvement has been exaggerated.

The ACER Report comments on the "poor performance" of 10 year-olds on newspaper items in 1975 when "only 27 per cent were successful" and the "marked improvement" in 1980 when the proportion of "successful students was 71 per cent", an improvement which made a substantial contribution to the higher levels of performance of 10 year-olds on the total reading test- from 53 per cent to 71 per cent (12). The Report indicates that the format of the newspaper was identical on both occasions and that items were "similar" or common (p156-157).

It is difficult to believe that a nation-wide improvement in student newspaper comprehension and reference skills of the magnitude indicated above did, in fact, occur. One indication of the difference in performance which we do not dispute lies in the comparisons between performances in 1975 and 1980 on the three common newspaper items utilized, viz. the last three common items shown in Table 7.5 of the

report (p80). On the first two common items, which assessed reference skills, there was a small but non-significant difference favouring the 1975 sample of 10 year-olds. On the last common item, which assessed comprehension of short passage about a male dolphin, the percentage of 10 year-olds correct in 1980 (86 per cent) was significantly higher than in 1975 (80 per cent). Overall, there were no significant differences between the performance in 1975 and 1980 of students on the common reading items and on the common items in the newspaper sub-test. Yet, on the newspaper sub-test as a whole, we are told that there was a marked difference in performance. How can this be?

The only explanation offered by ACER was that considerably greater emphasis may have been given to newspaper reading and comprehension in schools throughout Australia since 1975. The ASSP Project was not, of course, designed to collect evidence relating to changes in teaching practice. We have no evidence against which to test the validity of this suggestion, but we would like to suggest that other explanations cannot be ruled out. For example, it may be that the differences in student reading and reference skills were small (as indicated by the link items), and that the differences in performance can be attributed to differences between the tests used, not the students tested.

In 1980, the 10 year-old reading test contained 27 items, of which six items assessed comprehension and reference skills using a simulated two page newspaper. In 1975, the 10 year-old reading test contained 29 items, of which the last twelve formed the newspaper sub-test. On the last three of these items, there was a decrease in response rate (between 10 and 13 per cent of students did not attempt the items) indicating, according to the 1976 report, that some students found this section of the reading test too long. While it is true that the format of the newspaper used in 1975 and 1980 was similar and that there were three items common to both sub-tests, the newspaper sub-tests did differ considerably in length. In addition, when we compare the three non-common items employed in 1980 with the nine non-common items in 1975, we begin to have reasons for suspecting that the sub-tests differed in difficulty as well as length. In our view, there are no items in the 1980 sub-test comparable in difficulty to items 21, 23, 24 and 27 used in 1975. If this is the case much of the difference in the percentage of 10 year-olds classified as "mastering" the newspaper sub-test in 1975 and 1980 can be attributed to differences in the length and difficulty of the tests.

For 10 year-olds in 1975 to be classified as being successful, they would have had to obtain a score of at least 10 out of 12, whereas students in 1980 would have had to score 5 out of 6. As the ACER report (p53-54) indicates, if a longer sub-test is used, fewer students are likely to be classified as competent. Thus students in 1980 were more likely to be classified as competent because the sub-test was short and was completed by more students, the probability of misclassification as competent was increased, and, in our view, the non-common items were, as a whole, easier than those employed in 1975. The ACER acknowleges that part of the claimed improvement between 1975 and 1980 can be attributed to differences between the sub-tests. However, there was some evidence of upward movement from the trial

testing.

Attempting to make absolute comparisons across time, even when common tests and items are used, is not without its pitfalls. The problems are considerably greater when non-common tests and items are used. In the APU project, attempts were made to calibrate items and tests using the Rasch model. However, as noted in Chapter One, Goldstein has argued that it is not possible to calibrate the difficulties of items via other items, whose difficulties, for the sake of argument, are assumed to remain constant.

SCHOOL AND COMMUNITY REACTIONS

In order that reactions within the community to the results of the ASSP program might be more directly tapped, questionnaires were sent to selected organizations and individuals approximately one month after the release of the results. In each state 20 schools were contacted and it was arranged that copies of three questionnaires would be sent to each school. One questionnaire was for the principal to complete, one was intended for teachers of the tested populations (ie. English or Mathematics teachers of 14 year-olds, or primary teachers of 10 year-olds), and the third was for the secretary or president of the school's parent group. Stamped addressed envelopes were included to facilitate returns. The 20 schools in each state comprised 5 primary and 5 secondary schools that had participated in ASSP, and 5 primary and 5 secondary that had not: 140 schools in all.

A fourth questionnaire was sent to a selection of community groups within each state, each considered to have a potential interest in the ASSP and its outcomes. These groups included such bodies as teachers' associations, principals' associations, independent schools' associations, English and Mathematics subject area associations, local chambers of industry, apprenticeship commissions or boards, trade union groups, parents and friends' associations. Each of these questionnaires was also accompanied by a stamped addressed, return envelope.

Copies of each of these questionnaires are attached in the Appendices. The responses to each of the four questionnaires will be presented and discussed separately below. In each case, the response rate was low and even then many questions were unanswered or no opinion was put forward. To some degree at least the poor response may reflect the nature and level of school and community response to the project. In any event, the results should be treated with caution.

Principals

Returns were received from 61 principals, 33 from ASSP participating schools and 28 from non-participating. Of the participators, 16 were primary and 17 secondary.

Questions 2 and 3 were aimed at assessing the degree of interest in the ASSP results, as evidenced by the occurrence of discussion or meetings. Only 10 respondents reported any discussion of the ASSP results during staff meetings, and these for the most part were brief events of 15 minutes or less. No principals reported any ASSP committees as having been formed, and only 4 reported an intention of ever forming such committees in the future.

Question 4 sought to assess whether the outside community had approached the school (ie. principal) in any way in relation to the ASSP results. Only one principal reported having been contacted.

These responses of the principals would suggest that the ASSP results had very little direct effect on schools, at least at the top-end administrative level, and at least during the month immediately following the results' release. Of course individual principals may have given thought to the results, and influences may occur later in time. But, the immediate, direct effects seem to have been minimal.

TABLE 1: PRINCIPAL AND TEACHER RESPONSES TO QUESTIONS ON THE CLARITY AND USEFULNESS OF ASSP

Question	Principals (n=61) Yes	Principals (n=61) No	Teachers (n=141) Yes	Teachers (n=141) No
Was it clear that ASSP tested only some aspects of literacy and numeracy?	38 (62)*	8 (13)	45 (32)	19 (14)
Were the ways in which the results should and should not be interpreted made clear?	34 (56)	6 (10)	30 (21)	27 (19)
Were the national ASSP results found useful?	27 (44)	18 (30)	59 (42)	50 (36)
Is some sort of national testing program, perhaps different from ASSP, useful?	31 (51)	18 (29)	61 (43)	37 (26)

*Note: Figures in brackets are as a percentage of number of principal or teacher questionnaires returned.

Questions 5, 6 and 7 related to the reporting of the ASSP results. As can be seen from Table 1, the majority of principals felt the reporting of the ASSP results to be clear with respect to both

what was tested and how the results could be interpreted. However, the respondents did not seem so positive in judging the overall usefulness of either the ASSP in particular, or national testing in general. With regard to both these questions a sizeable proportion of respondents gave an explicit negative. The reasons given for "NO" responses to Questions 6 and 7 included claims that teacher union opposition had made the ASSP unworkable, indications that naturally occurring variations between schools made comparative interpretations difficult, that individual school results had arrived too late to be of use. Reasons for "YES" responses tended to focus around perceived benefits of benchmarks in setting and maintaining standards.

Questions 8 and 9 sought to determine the sources of the principals' information. The majority of principals first heard of the national ASSP results through the public news media (58 per cent) rather than via departmental (7 per cent) or ACER correspondence (20 per cent). Official interpretative material which was received, and less than half the respondents reported receiving any, tended to originate from ACER rather than state departments. This need not be an indication of tardiness on the part of state departments; it needs to be remembered that the ACER had the prime responsibility for the ASSP, and thus they got first look at the data.

Question 10 (see Table 2) solicited the principals' personal reactions to the actual results. The judgments were quite clear. For both age groups tested and for each of the performance categories, all but one or two principals felt that performances had been satisfactory. The interesting aspect of responses to this question is that less than 50 per cent of the respondents ventured an opinion. It would seem then that while those who have made a judgment are quite clear and united, there exists nonetheless some equivocation.

TABLE 2: OPINIONS OF PERFORMANCE ON ASSP (AS A PERCENTAGE OF TOTAL RESPONDENTS IN EACH GROUP)

Opinion	Principals* (n=61)	Teachers* (n=141)	Parent Groups (n=45)	Community Groups (n=48)
10 year old performance				
Better than satisfactory	2	4	4	4
Satisfactory	31	26	31	27
Unsatisfactory	4	2	2	4
No opinion	65	68	63	65
14 year old performance				
Better than satisfactory	2	6	0	6
Satisfactory	44	31	24	19
Unsatisfactory	2	1	0	8
No opinion	52	62	76	67

*Note: Teachers and principals rated reading, writing and numeracy performances. Percentages shown are mean ratings.

Teachers

Completed questionnaires were received from 55 primary and 86 secondary teachers. The patterns of response were very similar for primary and secondary teachers, so the two groups will not be discussed separately.

In parallel to the principals, teachers reported very little official school activity resulting directly from the ASSP report. Of the 141 respondents, 109 (77 per cent) reported no discussions of the ASSP results at staff meetings, and only four respondents reported some sort of staff sub-committee activity. Only one teacher was approached by someone from outside the school, a parent, to discuss the ASSP results. From the teachers' viewpoint as well as the principals', then, it seems that the ASSP results had very little direct immediate effect.

Teachers differed from the principals in that the teachers were less prone to endorse the ASSP reporting as being clear with regard to aspects tested and to interpretation (see Table 1). When it came to judging the usefulness of ASSP and nationwide testing, however, teacher respondents again paralleled the principals; 61 teachers felt that ASSP and nationwide testing were useful, 37 teachers responded negatively, and 43 either had no opinion or did not respond.

Teachers tended to first hear of the national ASSP results via the public news media (36 per cent), in parallel with the principals, except a further significant news source now appears, namely the principals (24 per cent). It would seem that even if the national ASSP results did not warrant discussion at staff meetings or the establishment of sub-committee groups, it was nonetheless of sufficient import in a proportion of schools for principals to at least mention the news to their teachers, presumably individually and informally. With regard to official correspondence relating to the ASSP results, only some 20 teacher respondents reported sighting any. This likely reflects only that most correspondence is directed to principals.

In their overall judgements of the adequacy of the reported student performances, teachers parallel the principals. For both age groups of students and for each performance area "it is clearly judged, by almost all teacher respondents, that performances were satisfactory". However, teachers seemed to show a greater degree of reluctance to respond than did the principals, in that teachers expressing a judgement on this question were an even smaller proportion (32 to 36 per cent) of the total than was the case for the principals.

Parent Groups

Completed questionnaires were received from 45 parent groups. Responses to Question 1 indicated that only a little over half the sample admitted to having any knowledge of the ASSP results. Of these 25 respondents, 12 learned of the ASSP results through the public news

media and 4 learned of them from their schools. The findings perhaps reflect two things; first, the very low level coverage given the ASSP report in the media, and second, the low level of immediate reaction by schools to the report as evidenced by the principals' and teachers' questionnaires discussed above.

Responses to Question 2 indicated that, in parallel with principals and teachers, parent groups exhibited very little immediate reaction to the ASSP report. The ASSP report was apparently only an infrequent topic of discussion, even in informal situations.

Parent group respondents did not tend to endorse the reporting of the ASSP results as being clear as to what was tested and as to what interpretations were justified. A minority of these groups indicated that both these aspects of the reporting were acceptable (see Table 3).

TABLE 3: PARENT AND COMMUNITY GROUP RESPONSES TO QUESTIONS ABOUT THE CLARITY AND USEFULNESS OF ASSP

Question	Parent (n=45)		Community (n=48)	
	Yes	No	Yes	No
Was it clear that ASSP tested only some aspects of literacy and numeracy?	13 (29)*	15 (33)	26 (54)	9 (18)
Were the ways in which the results should and should not be interpreted made clear?	5 (11)	18 (40)	13 (27)	19 (40)
Is some sort of nation-wide testing, perhaps different from ASSP, useful?	26 (58)	6 (13)	10 (21)	5 (10)

*Note: Figures in brackets are as a percentage of number of parent or community groups returning questionnaires.

Questions 4 and 5 attempted to tap into the respondents' opinions on national testing. Here, the majority of both the sample and those responding to the question (see Table 3) seemed to feel that nationwide testing was useful. These respondents looked to nationwide testing to inform on the levels of basic skill knowledge in children, to allow for comparisons between years, states, and schools (even though interstate and interschool comparisons are precluded), and to provide for problem detection. As to the results of the present ASSP project, those responding to the question clearly felt that performances at both age levels had been satisfactory (see Table 2). However, it needs be borne in mind that only a minority of the respondents actually ventured an opinion of the results.

Community Groups

Completed questionnaires were received from 48 community groups. Only eight of these respondents reported not having been aware of the release of the national ASSP results. Of the remaining 40, 34 reported first hearing of the results via the public news media.

The effect of the ASSP report's release on community groups appears to have been somewhat minimal, in parallel to its effect on the other groups surveyed. Only six respondents reported the ASSP results as having been discussed within their organizations, and only seven felt that any future discussion was likely.

With regard to the reporting of the ASSP results (see Table 3), it would seem that most community groups understood that the ASSP testing was confined. However, respondents tended to indicate that what constituted reasonable interpretations of the results was not properly understood. Those respondents venturing an opinion, favoured the retention of some sort of nationwide testing program. However, over half of the respondents explicitly indicated that their organizations had no official opinion on whether to continue such testing.

The minimal effect of the ASSP report on discussions within the responding community groups was also reflected in the virtual absence of any report-related contacts outside of the groups. Only five respondents reported contacting state education officials, and nobody reported contacting any government or parliamentary members. Further, only three indicated that such approaches were intended in the future, and these were likely to be directed only to schools.

Respondents seemed to indicate (see Table 2) that the levels of performance reported from ASSP were satisfactory, but only about a third of the respondents actually expressed an opinion; many of those that did not, wrote that they doubted their own expertise to form such an opinion. The National Council of the Australian Reading Association saw fit to publish a position paper on national testing (13). The Council questions the assumption that reading is being assessed by the ASSP tests since the tests do not focus directly on the reading process. It believes that the school should be the focal point of efforts to improve reading.

CONCLUSION

In all sectors of the community surveyed, the response to the results of the 1980 testing has been minimal in comparison with the 1975 results. At the political and media level, the results were noted briefly and put aside. One can only wonder if the reaction would have been as low key if the results had shown declines in a number of areas instead of an improvement. As it was, the feared resurgence of political attack on schools was not forthcoming; instead, the results have been used by the Ministers of Education to reassure the public, and occasionally by the ATF and the Opposition, in criticisms of the Fraser government. While the ATF seems to have shifted its position with respect to ASSP, the Education Departments contacted remain unenthusiastic and the VTU is still strongly opposed to ASSP. Few, if any, state policy or curriculum changes seem likely to be set in motion as a consequence of the release of the national results.

One major conclusion can be drawn from the questionnaire data. Very little discussion or activity took place in the month immediately following the release of the results. Further, from the reports of respondent principals, teachers, parents, and community group representatives, it seems that very little discussion or activity is expected in the future.

While overall it seemed understood that ASSP was confined to a specified range of literacy and numeracy skills, the limitations that this placed on interpretation of the results was not apparently clear. Only the principals of schools felt that interpretive limits were understood and made plain. The parent groups responding were very wary, a sizeable number indicating that it was not clear as to how the results should be interpreted.

With regard to the retention of nationwide testing, of some form or another, a majority in each of the groups surveyed (an overall average of two thirds) responded affirmatively and a minority (one third) negatively. However, not all groups responded equally strongly. Interestingly the community groups responded very tentatively indeed, with the majority stating that they had no official opinion.

On the performances reported in the ASSP results, the overall verdict seems to be that they were satisfactory. Again however, many respondents were nonetheless loathe to make a judgment. In no instance did more than half the respondents indicate an opinion.

NOTES AND REFERENCES

1. Bourke, S.F., Mills, J.M., Stanyon, J. & Holyer, F. Performance in Literacy and Numeracy: 1980 Canberra: Australian Government Publishing Service, 1981.

2. ibid., p.161.

3. AEC Press release, 9th September, 1981. Statement by Chairman of AEC, Mr. W.A.M. Gunn.

4. AEC Press Release. Summary of 1980 Report "National survey of basic skills".

5. Australia, House of Representatives, Debates Weekly Hansard No 19, 1981 p.2405.

6. The Bulletin, October 6, 1981.

7. The Canberra Times, October 29, 1981.

8. The Australian, September 21, 1981.

9. Inside Education, South Australian Education Department, 1981, 4(8).

10. South Australian Teachers Journal 1982, 14(1), p13.

11. Letter from General Secretary, Victorian Teachers Union in response to evaluation team's questionnaire.

12. Bourke, S.F., op. cit. pp. 83,86,87,156,157,161.

13. National testing in reading: a position paper. Australian Journal of Reading, 1982, 5(1), 47-48.

CHAPTER EIGHT

CONCLUSIONS

Colin Power

INTRODUCTION

> You cannot choose your battlefield,
> The Gods do that for you,
> But you can plant a standard
> Where a standard never flew.
>
> Nathalia Crane, The Colours.

Battles over standards of student behaviour and achievement have been forced on education systems on numerous occasions in the last one hundred years. In Australia, as in the US and the UK, the political response to the latest conflict has been to establish a national testing program. The AEC, with the assistance of the ACER, has planted ASSP in a battlefield where a standard never flew.

The evaluation team has seen its task as seeking evidence on questions raised by teachers, parents, interest groups and policy makers about ASSP, its purpose, origins, technical adequacy, implementation, and impact. It is also aware that the AEC does need to decide on the form and frequency of any future attempts to monitor levels of student performance. It has not been within our brief to make recommendations to the AEC, but it is hoped that the report will assist policy makers in assessing the advantages and disadvantages of national assessment.

PURPOSES OF ASSP

Were the purposes of ASSP as stated by the AEC achieved? ASSP was designed as an information-gathering exercise. Its purpose was to provide public information about the performance of 10 and 14 year-old students in basic skills areas at a national level. National estimates of the percentage of students achieving a minimum level of mastery in aspects of reading, writing and numeration were obtained. It was intended also that the attention of educators would be directed to areas in need of improvement or further investigation. The survey did reveal two areas (time and basic computation skills) in the performance of 10 year-olds which the AEC believes need attention, and

some areas (sex differences, patterns of errors in reading) in need of further research.

Generally then, the major purposes of ASSP were achieved. However, as the evidence presented here indicates, the information value of the results and the level of public knowledge of them can be questioned.

ASSP AND ACCOUNTABILITY

The notion of political accountability makes sense only if one is accountable to someone who has the capacity to inflict punishment. In 1976, the education system was called into account by the Commonwealth government and the media, and since then has had its funding cut. There seems to be little slackening in Australia, the US or the UK of the accountability movement. In the present climate, attempts to destroy political accountability systems are likely to be regarded as a confession of guilt. Currently, state education systems are being held politically accountable to governments for the results of their students on basic skills tests. But is this a just and appropriate accountability mechanism?

To the degree that educators are morally responsible for what they do to their students and with taxpayers money, they can and should be held accountable. Clearly, they ought not to be held responsible for outcomes not under their control (eg. youth unemployment). They may be called into account if all parties concerned are agreed that there is a body of teachable items which one can reasonably expect to be learned by the population of the school or system being judged; if the standards of quality and criteria of success are clear and agreed upon; if effective methods of teaching these items are known; and if the necessary resources have been provided. At the moment, these conditions are only partially met. For what, to whom and by what means should education systems and schools be held accountable, therefore, remains an unresolved issue. It is an issue, however, over which frank and open discussion among all parties involved is long overdue.

It may be that community concerns over standards and political accountability demands are a consequence, in part, of the lack of public understanding of what is happening in schools and the reasons for changes introduced. Systematic studies of changes in the content, processes and outcomes of schooling at the state and national level, (1) and improved dissemination procedures are needed (2). At the school level, the call by CDC for monitoring procedures with an agreed reporting mechanism acceptable to the community at large deserves more widespread attention than it has thus far received (3).

MANAGEMENT OF ASSP

ASSP was sponsored by the AEC and undertaken by the ACER. Recommendations on policy and the testing program came to the AEC through its Standing Committee which set up a Steering Committee and a

Technical Sub-Committee for the project. In the development of ASSP, the interests and the pooled expertise of the State Education Departments, ACER and ERDC were thus built into the management structure. Given the political context, these Committees faced a difficult task in formulating policy recommendations and guidelines for test development. This, in turn, created difficulties for the ACER in carrying out its task. The management structure has been criticised by ACER and by interest groups not represented; (e.g., teachers' unions, parent's organizations, non-government systems). Nevertheless, the committees and the ACER functioned, the program went ahead smoothly and the safeguards introduced were adhered to. In general then, we believe the management structure was sound. In retrospect, however, some of the difficulties may have been averted if there had been a more open debate and explanation of the proposed program prior to the AEC policy decision, if the guidelines given initially were clearer, if the management structure had been more representative, if there had been more consultation with professional groups with special expertise in educational measurement and the areas assessed, and more time between the decision to proceed and the date of testing.

THE TESTING PROGRAM

The tests developed by the ACER are criterion-referenced tests designed to measure aspects of student performance in reading, writing and numeration. Among the major political parties there would appear to be agreement on the importance of literacy and numeracy , even if there are differences in how this concern is expressed. Surveys of community expectations of schooling show that all groups in the community believe the skills ASSP seeks to assess are extremely important (4). Agreement was also reached within the AEC Steering Committee for ASSP and its Technical Sub-Committee about the importance of the specific areas assessed, although at times, alternative views of what it means to be literate and numerate (cf. the essentialist, interactionist, instrumentalist, and conflict models) have been injected into the debate. Within the limits of the existing technology of criterion-referenced testing, the testing program was carefully developed, and the scales and tests used were reliable (generally 0.8 or above).

The measurement literature (5) indicates that there are unresolved technical problems in establishing construct validity when criterion-referenced tests are used to assess skills like reading and writing, and when they are used to evaluate educational programs. The validity of tests like those used in ASSP depends on whether the purposes of the tests and the areas to be assessed can be clearly specified, on the adequacy of the item selection process, and the degree to which cut-off scores can be justified. It is difficult to establish the validity of tests if the purpose is to ensure young people have the skills necessary for success in adult life. The development of tests of school-related skills is less problematic, but it is not easy to define competence in reading, writing and numeration clearly enough to make item selection a straightforward process. This problem is compounded when the task involves developing tests with

items all with difficulty levels focussed at the point of minimum competence. As a consequence, conclusions about performance standards are not independent of the difficulty of the sample of items selected.

The establishment of cut-off points is one of the most controversial problems in criterion-referenced testing. All standard setting techniques are arbitary and fallible, and further evidence of the validity of the points chosen is needed. There is some evidence (6) that the level chosen in 1975 is related to the tendency to stay on at school, and, for numeracy but not literacy, to employment prospects. But whether one is assessing literacy, numeracy, success in later schooling or life, it is difficult to select a single cut-off such that "mastery" implies successful performance in the task domain of interest and unsuccessful performance does not. The difficulty is exacerbated when another layer of standards is imposed to determine if an acceptable proportion of students throughout the nation is answering an acceptable number of items.

The alternative to assuming mastery as an all or nothing affair, is to focus on the whole performance continuum. In our view, assessing over the whole range of performance in reading, writing and numeracy is likely to give a more valuable return for investment than focussing on one level only.

COSTS

The expenditure incurred by ACER on the 1980 ASSP project was $235,543, or $22.27 per student tested. The ERDC grant to the evaluation team was $15,620. To these amounts must be added the travel and other expenses incurred by State Department Officers attending Steering and Technical Committee meetings, and the hidden costs in terms of time of the school personnel involved. These costs are modest in comparison to larger national assessment projects like APU and NAEP.

IMPLEMENTATION OF THE PROGRAM

The boycott by Teachers Unions created difficulties for the ACER in implementing the program. Although they had agreed to participate in the testing program, 139 schools withdrew from the study as a result of teachers' union or parent association pressure so late that it was impossible to replace them.

In the participating schools, the implementation of the testing program was carried out efficiently with a minimum of fuss and with the general support of parents and most staff. However, about one quarter of the teachers were unconvinced of the value of the testing or suspicious of the political uses to which the data might be put.

Schools requesting results for their own students received them as promised, and all participating schools received the summary report of the national results once this became available. As the intention was to inform educators and the public, consideration might have been

given to distributing the summary report to all schools and major interest groups.

ESTABLISHING NATIONAL ESTIMATES

The details of the testing program, sampling and results are contained in the AEC report of the project prepared by the ACER. The critical question to be asked here is whether the national estimates are sound and the conclusions reached are justifiable the light of the evidence obtained. The ACER has avoided making judgements as to whether performance levels reported are satisfactory, there being no objective basis for doing so. It was able in most cases to reach defensible conclusions about changes in performance between 1975 and 1980 and differences between sexes and age groups. In reaching these conclusions about differences between 1975 and 1980, care was taken to consider the effects on national estimates of differences in the samples employed on such factors as school location, year level, age and language spoken at home, and of the withdrawal of schools. Analyses undertaken by the evaluation team indicate that the relatively high non-response rate in some sub-samples does not appear to have unduly biased the national estimates.

IMPACT ON THE PRACTICES OF PARTICIPATING SCHOOLS

ASSP has had no appreciable impact on the teaching practices, curriculum or assessment procedures of the participating schools studied. The national character of the program and the safeguards employed (light sampling, confidentiality etc.) seem to have been effective in preventing any adverse backwash effects on the schools involved. There was no evidence of coaching for the tests. Only one school made a change to the teaching of basic skills on the basis of the results received. Not all schools looked at their results. Schools doing so, indicated that the results were as expected or higher than expected. While almost all participating schools were prepared to continue to cooperate with the national testing project, they indicated a higher level of support for a school-based literacy and numeracy testing program.

IMPACT OF THE NATIONAL RESULTS

Politically, the impact of the national results has been limited. The results were welcomed by the AEC and used in an attempt to reassure the public about standards. However, the press releases of the AEC were given little publicity by the media. Those who attacked schools over standards in basic skills offered no comment, and those concerned about academic standards dismissed the results as irrelevant. The ATF and the opposition made some limited use of the data in political attacks on the Fraser government and in defence of the school system.

At the systems level, some states plan to analyse data tapes for their state, summaries of the results have appeared in official publications, and the report has been made available to appropriate curriculum committees. Nevertheless, the State Education Departments have sought to keep ASSP in low profile. Given the very limited scope of the testing, the results have limited utility for policy making or program evaluation at the state level. Thus far, the impact on the policy and programs of the various systems systems has been small.

At the school level, there was little discussion of the results either by teachers or parent organizations. The schools surveyed are not planning to make significant changes in emphasis, materials or methods. There seems to be no compelling reason for them to do so. Firstly, the results are at a national, not a school, level. The evidence from this study, the US (7) and the UK (8) suggests that teachers make little use of the results of national assessment. They are much more likely to make changes in their programs and initiate remedial work with students lacking in basic skills on the basis of diagnostic information from school-based testing programs. For this reason most educators (including the Evaluation Team) strongly support the efforts being make to develop appropriate forms of testing for school-based diagnosis and remediation as part of the Australian Co-operative Assessment Project. Only one half of school parent organizations surveyed admitted to having any knowledge of the results, and many of these were unclear as to how the results should be interpreted. Where an opinion was ventured, the level of performance of students was judged to be satisfactory. In few cases was there any discussion on ASSP between teachers and parent groups in schools. The state and national community and educational organizations surveyed were more aware of the results, but again many groups were unsure of their meaning and limitations.

Of the teachers, principals, parent groups and community organizations venturing an opinion about the desirability of a national monitoring program in some form or other, two-thirds supported the notion, and one-third did not.

HOPES AND FEARS

Supporters of national testing hoped that the results would:

(a) provide unassailable evidence about changes in standards which would settle disputes on the issue, and/or
(b) lead to action on the part of policy makers, education systems, research workers, curriculum developers and schools which would ultimately result in fewer students leaving school deficient in "basic" skills.

The degree to which these benefits can be expected to eventuate is limited. True, some heat has been taken out of the standards debate. But the standards issue is political; the grounds for debate shift; and the evidence is as likely to be misused or dismissed as it is to settle arguments. ASSP does not provide evidence of standards other than at one arbitrary, minimal level in selected basic skills. No

information is available about other levels of performance, other subject areas or standards of behaviour. No information is available about changes in the curriculum, teaching methods or family circumstances which might account for the shifts in performance observed. Fears have been allayed in one area, but not in others which also are pertinent to questions about the quality of education in Australian schools.

It seems fairly unlikely that as a direct result of ASSP, fewer students will leave school deficient in the areas of weakness identified in the 1980 testing. There is little evidence that the results have led to significant changes in policy or practice at any level (national, state, system, school or classroom). For appropriate action to be taken, the data must be pertinent to the decisions being made and the decision-making level. Moreover, one must understand what has caused differences, rises and falls in achievement; and one must know what, if anything, can be done to overcome problems identified. The prospects for following-up the 1980 testing results with research directed at the questions raised by the report are not good. ERDC has been abolished, research budgets in tertiary institutions and State Departments are being pruned, and no investigations on the issues raised on the report have been commissioned. Without an associated research and development program aimed at contributing to an analysis of the problems of illiteracy and innumeracy in our society and to the informing of policy and practice, the information value of ASSP is severely limited.

If the hopes of supportors of ASSP are unlikely to be realised, the fears of its opponents also seem equally unlikely to eventuate. There was no evidence of an adverse backwash effect on teaching methods or on curriculum balance. The feared political misuses of the data and attack on schools also did not take place, in part because of the work of the AEC Steering Committee, and in part because the results showed no serious declines in any area tested.

WIDER PROSPECTS

The absence of any immediate effect does not necessarily mean that there will be no impact in the long run. By repeatedly drawing attention to particular skills and minimum standards, ASSP may lead the community to begin to judge the effectiveness of schools in terms of those skills and standards. Eventually, a greater share of the resources in time and personnel may come to be allocated to those areas and students seen to be at risk, but less to average and above average students and to striving for excellence in these and other areas for which schools are responsible. Moreover, it may take some time (at least 4 to 5 years) for the results of national testing to have an impact on state systems and individual schools. The problems identified by the 1980 testing are long standing ones, but renewed efforts to solve them may ultimately lead to identifiable improvements in areas of weakness. Given that it will take time for the results of responses to a given testing to show up, serious questions must be asked about the desirability of annual testing. In our view, if national testing were to be undertaken annually, the costs (both

financial and educational) would far outweigh any benefits likely to accrue.

CONCLUDING COMMENTS

The evidence presented in this report suggests that it has not been easy to create and plant a standard which all understand and all will accept. Some progress has been made, but much remains to be done. It is our hope that as a consequence of this evaluation, educators and other interested people will be in a better position to appreciate what has been achieved, and to understand the nature of the difficulties, political and technical, which have been, and still have to be, faced in the development of a national system for monitoring student performance.

We have expressed concern that many in the community with a stake in the outcomes of schooling either are not aware of, or do not understand, why ASSP was introduced and what inferences about schooling can legitimately be made from the results. Moreover, it needs to be understood that consensus over the value of ASSP is unlikely to be reached (indeed it has not been within the evaluation team). The strongly held views of supporters and opponents of ASSP reflect fundamental differences in beliefs and values. Supporters of ASSP place great value on literacy and numeracy, believe that schools do not stress these enough, and that, in the absence of an accountability mechanism like ASSP, standards of literacy and numeracy are likely to suffer. Opponents of ASSP value a broader educational experience, believe that sufficient emphasis is being given to the development of literacy and numeracy skills, and that there are better ways of monitoring and improving the quality of education. Now that the testing is over, some shifts in viewpoint about the appropriateness of ASSP as a means of informing the public and improving the quality of education may have taken place. Given the restricted scope of the testing, the limited information value and impact of the results and the inadequacies of the media coverage, some will wonder whether there is really much point in repeating the exercise. Some will use the evidence presented here to reassess the strengths and limitations of national testing, others will cling to their hopes, and still others to their fears.

Given the divided opinions about the value of national assessment and the current state of public knowledge, it will be necessary for the AEC to do more than announce its intentions regarding the future of national assessment in Australia. It will need to explain carefully its position and the reasons for its decision. It will also need to seek to ensure that the public receives a balanced account of the issues addressed and the evidence considered in reaching a verdict.

NOTES AND REFERENCES

1. Rosier, M. Changes in secondary school mathematics in Australia 1964-78. Melbourne: Australian Council for Educational Research, 1980.

2. Power, C.N. The contribution of research to educational policy and practice. Australian Educational Researcher, March 1981, 5-19. The thrust in ERDC prior to its abolition was along the lines being argued for here.

3. CDC. Core curriculum for Australian schools. Canberra, CDC, 1980.

4. Bourke, S. Community expectations of numeracy in schools Australian Journal of Education, 1979, 23(2), 142-152.

Baumgart, N. Survey of public opinions about schools in New South Wales. Sydney: Macquarie University, 1979.

Mason, K. An evaluation of research on community expectations of education. Canberra: Canberra Colege of Advanced Education 1979.

5. Shepard, L. Technical issues in minimum competency testing. In D.C. Berliner (ed.), Review of research in Education Vol.8. Washington, American Education Research Association, 1980.

Bunda, M.A. & Sanders, J.R. Practices and problems in competency-based measurement. Washington: National Council on Measurement in Education, 1979.

6. Williams, T., et al. School, work and career. Melbourne: ACER, 1980, p96.

7. Resnick, L.B. Introduction: research to inform a debate. Phi Delta Kappan, May, 1981 p.623.

8. Wood, R. & Gipps, C. An inquiry into the use of test results for accountability purposes. In Calling education into account. Open University Course E364, Heinemann (in press).

APPENDICES

		Page
1.	Pre-testing	126
2.	Implementation questionnaires	128
3.	Questionnaires: Initial impact of ASSP results	134
4.	Questionnaires: Reaction to ASSP post-results	142

EVALUATION OF AUSTRALIAN STUDIES IN

STUDENT PERFORMANCE

NAME OF SCHOOL: .....

TYPE OF SCHOOL: ☐ STATE ☐ CATHOLIC ☐ INDEPENDENT

LOCATION: ☐ METROPOLITAN ☐ PROVINCIAL CITY (>10,000) ☐ RURAL

ENROLMENT: _____ pupils

- -

1. Were you aware that a national survey of the performance of 10 and 14 year olds in aspects of reading, writing and numeracy is about to be undertaken? (Tick one box)

YES ☐ NO ☐

2. Has your school been invited by the A.C.E.R. to participate in the A.S.S.P. project?

YES ☐ NO ☐

(If your answer is NO, you need not answer any further questions)

3. Has your school agreed to participate in the project?

YES ☐ NO ☐

Please indicate the reason(s) for your decision:

.....

.....

.....

4. Would your school be willing to assist in the evaluation of the A.S.S.P. project?

YES ☐ NO ☐

5. Did the material your school received from A.C.E.R. about the project tell you all you needed to know to make a decision about participation?

YES ☐ NO ☐

If NO indicate what additional information you would like to have received:

.....

.....

.....

.....

.....

6. Who at your school was involved in reaching a decision about participation in the project? (Tick appropriate box or boxes)

- ☐ Principal
- ☐ Teachers of classes concerned
- ☐ Parents of children concerned
- ☐ Senior Staff
- ☐ Whole Staff
- ☐ Parents organization or School Council

7. How was the decision reached?

- ☐ Decision treated as an executive matter to be handled by principal.
- ☐ Principal made decision after consulting with staff and/or parents organization.
- ☐ Decision was the result of a vote taken or consensus reached at meeting of staff and/or parents.

8. Has request by A.C.E.R. for your school to participate in the project had any effect on policy, curriculum, teaching or evaluation procedures in your school?

YES ☐ NO ☐

If YES, please give details:

.....

.....

.....

.....

CONFIDENTIAL

EVALUATION OF AUSTRALIAN STUDIES IN STUDENT PERFORMANCE

IMPLEMENTATION OF ASSP, QUESTIONNAIRE (1) - PRINCIPALS/ADMINISTRATORS

NAME OF SCHOOL: .....

---

1. Were communications from the A.C.E.R. (information, instructions, test materials, etc) associated with the implementation of the survey of reading, writing and numeracy received in good time?

YES ☐ NO ☐

If NO, what were the difficulties?

.....

.....

.....

2. Have the instructions received from the A.C.E.R., regarding the implementation of the survey, been clear and adequate on:

(a) the sampling of pupils? YES ☐ NO ☐

(b) the administration of the tests? YES ☐ NO ☐

(c) the collection and return of used and unused test materials? YES ☐ NO ☐

If you have answered NO to either (a), (b) or (c) please describe the deficiency, and any action taken to correct it.

.....

.....

.....

3. Have there been any difficulties in your school in complying with the instructions of the A.C.E.R. on:

(a) the sampling of pupils? YES ☐ NO ☐

(b) the administration of the tests? YES ☐ NO ☐

(c) the collection and reutrn of used and unused test materials? YES ☐ NO ☐

If you have answered YES to either (a), (b) or (c) please describe the difficulty and how you responded to it.

.....

.....

.....

P.T.O.

4. (a) When were the pupils made aware that they were to be tested?

.....

(b) What was the source of this information and how was it presented?

.....

.....

(c) What were they told about the testing?

.....

.....

.....

5. Did the awareness of the testing programme affect the pupils to be tested and/or their teachers in any way?

YES ☐ _ NO ☐

If NO, why do you think pupils/teachers were unaffected?

If YES, please describe the nature of the effects?

(a) pupils .....

.....

(b) teachers..... .....

.....

6. Please give an estimate of the number of man hours expended by school staff on the implementation of the survey

(a) Number of man hours spent in testing ☐

(b) Number of man hours spent in other activities associated with the implementation ☐

(c) What was the nature of these 'other activities'?

.....

.....

7. (a) What effects, beneficial or harmful, do you believe the survey will have for your school?

.....

.....

.....

(b) What effects, beneficial or harmful, do you believe that survey will have beyond the school?

.....

.....

.....

6/10/80
JT.

EVALUATION OF AUSTRALIAN STUDIES IN STUDENT PERFORMANCE

IMPLEMENTATION OF ASSP, QUESTIONNAIRE (2) - TEACHERS

NAME OF SCHOOL .....

---

1. Of the children you teach, how many have been tested in the national survey of reading, writing, and numeracy? From how many classes were these children drawn?

   Number of children ☐ Number of classes ☐

2. At which level do you teach? Primary ☐ Secondary ☐

   If SECONDARY, which of the following subjects do you teach these children:

   English? YES ☐ NO ☐

   Mathematics? YES ☐ NO ☐

   Other subjects? (please name) .....

3. How did you first learn about the nature and purposes of the survey?

   ☐ By reading the printed material distributed by A.C.E.R.

   ☐ Through discussion with other members of staff

   ☐ A ministerial statement or press release

   ☐ Through other channels (please state).....

   .....

4. Did the information you received from A.C.E.R adequately explain the nature and purposes of the survey?

   YES ☐ NO ☐

   If NO, what further information would you have liked?

   .....

   .....

   .....

5. Did you have the opportunity of contributing to the decision of the school to participate in the survey?

   YES ☐ NO ☐

   If YES, how were you able to contribute?

   .....

   .....

   .....

6. Do you believe that the school should be participating in the survey?

YES ☐ NO ☐

What are your main reasons for this belief?

.....

.....

.....

7. Prior to the testing, did you have any information on the nature of the tests or the kinds of test questions that were to be given?

YES ☐ NO ☐

If YES, from what source did this information come? What use did you make of it?

.....

.....

.....

8. Has the fact that your school has participated in the project had any effect on policy, curriculum, teaching or evaluation procedures in your school?

YES ☐ NO ☐

If YES, please give details.

(a) policy .....

.....

(b) curriculum .....

.....

(c) teaching .....

.....

(d) evaluation procedures .....

.....

6/10/80
JT.

CONFIDENTIAL

EVALUATION OF AUSTRALIAN STUDIES IN STUDENT PERFORMANCE

IMPLEMENTATION OF ASSP, QUESTIONNAIRE (3) - PARENTS

NAME OF SCHOOL .....

---

1. Did you know that one of your children has been tested at school in a national survey of reading, writing and numeracy conducted by the Australian Council for Educational Research?

YES ☐ NO ☐

(If your answer is NO, you need not answer any further questions.)

2. What is the age of your child who is being tested in the national survey of reading, writing, and numeracy?

10 ☐ 14 ☐

3. Do you know something about the nature and the purposes of the survey?

YES ☐ NO ☐

If YES, how did you first learn about the nature and purposes of the survey? (Tick one or more boxes)

☐ By reading printed material sent out by the school

☐ Through a parents' meeting or a meeting of staff and parents

☐ Through other channels (please state)

.....

.....

4. Did the information you received adequately explain the nature and purposes of the survey?

YES ☐ NO ☐

If NO, what further information would you have liked?

.....

.....

.....

P.T.O.

5. Did you have the opportunity of contributing to the decision of the school to participate in the survey?

YES ☐ NO ☐

If YES, how were you able to contribute?

.....
.....
.....

6. Do you believe the school should be participating in the survey?

YES ☐ NO ☐

What are your main reasons for this belief?

.....
.....
.....

7. Has your child discussed the survey at home?

YES ☐ NO ☐

If YES, what has he/she said about the survey?

.....
.....
.....

Did he/she appear particularly concerned over the testing?

YES ☐ NO ☐

8. Have you been aware of any special preparation for the survey being carried out by the school or your child?

YES ☐ NO ☐

If YES, please give details.

.....
.....
.....

6/10/80
JT.

INITIAL IMPACT OF ASSP RESULTS ON SCHOOLS

PRINCIPALS' QUESTIONNAIRE

Interviewer meets principal by pre-arranged appointment. Interviewer explains about the team of researchers evaluating ASSP ... explains that this meeting is part of the follow-up evaluation using a selected group of schools as case-studies.

	ASKS THE PRINCIPAL ...	ENTERS THE RESPONSES ...
Q1	Which group of children in your school were tested in ASSP program?	10 yr olds ☐ 14 yr olds ☐
Q2	Did you request a print-out of your students' results?	Yes ☐ No ☐
	(if no, record why and then gently abandon the interview)	
Q3	Have you received the print-out?	Yes ☐ No ☐
	(if no, seek another appointment after recording Q4 response)	
Q4	Any delays or concerns in the mechanics of requesting and receiving the print-out?	
	.....	
	.....	
	.....	
Q5	How long have you had the print-out of your students' results?	up to one week ☐ about ten days ☐ two weeks or so ☐ three weeks or more ☐
Q6	Have you read through the print-out?	Yes ☐ No ☐
Q7	Were there any problems in understanding it or in the way it was set out?	Yes ☐ No ☐
	.....	
	.....	
	.....	
Q8	Did you try and analyse the significance of the results for your school and teaching staff?	Yes ☐ No ☐ (not really)

IF YES

Q9 (a) What significant factors emerged?

.....

.....

.....

(b) What are the main implications for your teachers?

.....

.....

.....

(c) What are the main implications for your school's curriculum?

.....

.....

.....

Q10 To what extent do you agree with the (print-out) results of your students in each of the areas tested?

	To a great extent	To a moderate extent	To little or no extent
Reading			
Writing			
Numeration			

Comment if any: .....

.....

.....

.....

Q11 To what extent do you feel satisfied with the results which your students obtained in each of the areas tested?

	To a great extent	To a moderate extent	To little or no extent
Reading			
Writing			
Numeration			

Comment if any: .....

.....

.....

.....

Q12 Have you ensured that the students' results will not be publicly divulged? Yes ☐ No ☐

Any comment on this aspect: .....

.....

.....

.....

– 3 –

Q13 To what extent have you been subjected to "pressure" to divulge your students' results?

Degree of Pressure **by** **Which Individuals/Groups**

.....

.....

Q14 Have you consulted with any of the following on the results shown in the print-out?

	Yes	No	Not Applicable
(a) Superordinates (Specify which .....)			
(b) Deputy principal			
(c) Teachers of participating pupils			
(d) Other teachers of the year group of participating pupils			
(e) Other English teachers			
(f) Other Maths teachers			
(g) Other staff members			
(h) General staff meeting			
(i) Other consultancies: Specify who			

Q15 Have you as yet told any of your staff to do anything as a result of your analysis of the print-out or your consultations? Yes ☐ No ☐

WHO EXACTLY has been told to do WHAT?

.....

.....

.....

.....

.....

(these names and jobs are important as the interviewer is to conduct follow-through interviews with nominated persons)

Q16 Have you yourself initiated any other moves to enable something to be done as a result of your analysis of the print-out or your consultation with staff? Yes ☐ No ☐

WHAT MOVES INITIATED to enable WHAT TO BE DONE

.....

.....

.....

.....

.....

Q17 (Resulting from Q15 and/or Q16) Has anything actually been done as yet? Yes ☐ No ☐

WHAT EXACTLY (See also Q18 (a), (b) and (c))

.....

.....

.....

.....

Q18 Have the ASSP results had any impact on

(a) lesson time allocated to

	Yes	No
Reading		
Writing		
Numeration		
Other Subjects		

Please describe the change: .....

.....

.....

(b) strategies used in the teaching of

	Yes	No
Reading		
Writing		
Numeration		
Other Subjects		

Please describe the change: .....

.....

.....

(c) remediation activities in

	Yes	No
Reading		
Writing		
Numeration		
Other Subjects		

Please describe the change: .....

.....

.....

Q19 Do you have any firm intention of making changes in the near future which you would say have directly resulted from the students' participation in the ASSP project or their scores on the tests?

Yes ☐
No ☐

Describe intended changes: .....

.....

.....

.....

Q20 Are there any further comments on ASSP which you would like to make?

.....

.....

.....

.....

.....

THANK YOU VERY MUCH FOR YOUR TIME AND HELP

*(Do you think you could now help me to arrange an interview to see .....
(person(s) nominated in Q14 and/or Q15))*

INITIAL IMPACT OF ASSP RESULTS ON SCHOOLS

SCHOOL STAFF QUESTIONNAIRE

Interviewer meets staff member via the principal's nomination as a person consulted (Q14 of the Principal's Questionnaire*) or a person told to do something (Q15 of the* Principal's Questionnaire*). Interview arranged with the principal's help.*

Interviewer specifies the position of the interviewee (Deputy Principal, Year 6 class teacher, etc.)

Interviewee: .....

Interviewer explains that this meeting is part of a follow-up evaluation of the ASSP project, using a selected group of schools as case studies.

Q1 Have you read through the print-out of the results of the students at your school? Yes ☐ No ☐

IF NO,

Q2 But the principal has consulted you about the results and indicated the trends in them? Yes ☐ No ☐

IF NO, GO TO Q6

Q3 To what extent do you agree with the (print-out) results of the students in this school in each of the areas tested?

	To a great extent	To a moderate extent	To little or no extent	Not Applicable
Reading				
Writing				
Number work				

Comment, if any: .....

.....

.....

.....

Q4 To what extent do you feel satisfied with the results which the students obtained in each of the areas tested?

	To a great extent	To a moderate extent	To little or no extent	Not Applicable
Reading				
Writing				
Number work				

Comment if any: .....

.....

.....

.....

Q5 Have you yourself made any moves as yet as a result of the ASSP results for the school? Yes ☐ No ☐

WHAT MOVES? and WHY?

.....

.....

.....

.....

.....

.....

Q6 Have the ASSP results had any impact on

(a) lesson time allocated to

	Yes	No
Reading		
Writing		
Numeration		

Please describe the changes: .....

.....

.....

.....

.....

(b) strategies used in the teaching of

	Yes	No
Reading		
Writing		
Numeration		

Please describe the changes: .....

.....

.....

.....

.....

(c) remediation activities in

	Yes	No
Reading		
Writing		
Numeration		

Please describe the changes: .....

.....

.....

.....

.....

Q7 Have the ASSP results had any other effects, *whether positive or negative*, on

Yes ☐
No ☐

(a) teachers in the school

Please explain: .....

.....

.....

(b) teaching methods

Please explain: .....

.....

.....

(c) curriculum

Please explain: .....

.....

.....

(d) other aspects of school life

Please specify and explain: .....

.....

.....

Q8 Do you have any firm intention of making changes in the near future, which you would say have directly resulted from the students' participation in the ASSP project or their scores on the tests?

Yes ☐
No ☐

Please describe intended changes: .....

.....

.....

.....

Q9 Are there any further comments on ASSP which you would like to make?

.....

.....

.....

THANK YOU VERY MUCH FOR YOUR HELP

Reactions to ASSP - Post-results

PRINCIPALS

Name of School _____

Date _____

1. Was your school involved in the recent ASSP testing of reading, writing and numeracy?

   Yes ☐ Age levels? 10 years ☐
   14 years ☐

   No ☐

2. Since the release of the ASSP results, have meetings of staff been called to discuss these results?

   No ☐

   Yes ☐ a. How many meetings? _____

   b. Approximate total duration of discussion on ASSP results

   15 mins. ( )
   30 mins. to 1 hour. ( )
   More than 1 hour. ( )

   Are such meetings planned for the near future?

   No ☐

   Yes ☐

3. Have sub-committees of staff been formed to discuss any possible implications that the ASSP results might have for curriculum modification in your school?

   No ☐

   Yes ☐ a. How many staff members are involved? _____

   b. What are the teaching areas of the staff involved? _____

   Are such sub-committees planned for the near future?

   No ☐

   Yes ☐

4. Since the release of the ASSP results, have you been approached in any way by outside parties to discuss the relevance of those results to your school?

   No ☐

   Yes ☐ a. Approximate number of contacts _____

   b. Please indicate:

   Parents ( )
   Business/Commercial ( )
   Other (Specify) ( )

5. In the reporting of the ASSP results by the ACER,

   (a) was the restricted range of the topics tested made clear?

   Yes ☐

   No ☐

   (b) were the interpretations that can and cannot be made of the results clearly explained?

   Yes ☐

   No ☐

   Comments _____

6. Do you find that the ASSP results are useful in any way?

   Yes ☐

   No ☐

   Please elaborate on your reasons _____

7. Do you think that some sort of national testing program is useful?

   Yes ☐ Please specify what sort of format you would like to see such a program take.

   _____

   _____

   _____

   No ☐ Reasons?

   _____

   _____

   _____

8. By what channel did you first hear of the ASSP results?

   TV/Radio/Newspapers ☐
   Correspondence from state education department ☐
   Report from A.C.E.R./A.S.S.P. ☐
   Other (Specify) ☐

   _____

9. Have you received any official information from either your state department or the ACER which outlines and interprets the results of the ASSP program?

Yes ☐ From whom?
State Department ☐
A.C.E.R. ☐

No ☐

10. What are your reactions to the results? In overall terms, were they:

		Reading	Writing	Numeracy
(a)	10 year olds			
	i. Satisfactory?	☐	☐	☐
	ii. Better than satisfactory?	☐	☐	☐
	iii. Below what would have been satisfactory?	☐	☐	☐
(b)	14 year olds			
	i. Satisfactory?	☐	☐	☐
	ii. Better than satisfactory?	☐	☐	☐
	iii. Below what would have been satisfactory?	☐	☐	☐

Reactions to ASSP - Post-results TEACHER

Name of School _____

Date _____

1. What are your teaching fields?

Primary ☐ Secondary ☐

English ☐
Mathematics ☐
Other ☐
----------(specify)

2. Approximately what percentage of the students that you teach were involved in the A.S.S.P. testing? Mark only the applicable group.

10 year olds _____%

14 year olds _____%

3. Have staff meetings been called, since the release of the A.S.S.P. results, to discuss those results?

No ☐
Yes ☐ a. How many meetings? _____

b. Approximate total duration of discussion(s) on A.S.S.P. results

15 mins. ( )
30 mins. to 1 hour. ( )
More than 1 hour. ( )

Are such meetings planned for the near future?

No ☐
Yes ☐

4. Have sub-committees of staff been formed to discuss any possible implications that the A.S.S.P. results might have for curriculum modification in your school?

No ☐
Yes ☐ a. How many staff members are involved? _____

b. What are the teaching areas of the staff members involved? _____

Are such subcommittees planned for the near future?

Yes ☐
No ☐

5. Since the release of the A.S.S.P. results, have you been approached in any way by outside parties to discuss the relevance of those results to your school?

No ☐
Yes ☐ a. Approximate number of contacts _____

b. Please list:

Parents ( )
Business/Commercial ( )
Other (Specify) ( ) _____

6. How would you assess the way in which the A.S.S.P. results were reported by the A.C.E.R.?

(a) Was the restricted range of the topics tested made clear? Yes ☐ No ☐ No opinion ☐

Any comments? _____

(b) Were the interpretations that can and cannot be made of the results clearly explained? Yes ☐ No ☐ No opinion ☐

Any comments? _____

7. Do you find that the A.S.S.P. results are useful in any way?

Yes ☐
No ☐

Please elaborate on your reasons for your answer _____

_____

8. Do you think that some sort of national testing programme is useful?

Yes ☐ Please specify what sort of format you would like to see such a programme take e.g. age levels tested, topics tested, frequency of testing, sampling employed.

_____

_____

_____

No ☐ Please give reasons e.g. what "alternative" approach might you find more useful and why?

_____

_____

_____

9. By what channel did you first hear of the A.S.S.P. results?

TV/Radio/Newspapers ☐
Correspondence from state education department ☐
Report from A.C.E.R./A.S.S.P. ☐
Other (Specify) ☐

---

10. Have you received any official information from either your state department or the A.C.E.R. which outlines and interprets the results of the A.S.S.P. programme?

Yes ☐ From whom?
(a) State department ☐
(b) A.C.E.R. ☐

No ☐

11. In your estimation, have the A.S.S.P. results been an issue of much discussion in your school's staff room?

Yes ☐ No ☐

12. What are your reactions to the results? In overall terms, were they:

	Reading	Writing	Numeracy
(a) 10 year olds			
i. Satisfactory?	☐	☐	☐
ii. Better than satisfactory?	☐	☐	☐
iii. Below what would have been satisfactory?	☐	☐	☐
(b) 14 year olds			
i. Satisfactory?	☐	☐	☐
ii. Better than satisfactory?	☐	☐	☐
iii. Below what would have been satisfactory?	☐	☐	☐

Reactions to ASSP - Post-results

PARENT

1. Have you heard any information on the results of the recent nationwide testing program on reading, writing, and numeracy?

   Yes ☐ By what means? .....

   No ☐

2. Was one of your children tested as part of this program?

   Yes ☐ His/her age? ..... years

   His/her school? .....

   No ☐

3. Have you discussed the testing results with

   .... your child's principal Yes ☐

   .... any of your child's teachers Yes ☐

   .... with other parents at the school Yes ☐

   .... with any of your friends or neighbours Yes ☐

4. In the reports that you have seen or heard,

   (a) was the restricted range of the topics tested made clear?

   Yes ☐ No ☐

   (b) were the ways in which the results can and cannot be interpreted made clear?

   Yes ☐ No ☐

5. Do you think that some sort of nationwide testing program is useful?

   Yes ☐ What sort? .....

   Why? .....

   No ☐ Why not? .....

6. For the 10 year olds, do you feel that the results were:

   (a) satisfactory? ☐

   (b) better than what you would have felt to be satisfactory? ☐

   (c) below what you would have felt to be satisfactory? ☐

6. (Continued)

   For the 14 year olds, do you feel that the results were:

   (a) satisfactory? ☐

   (b) better than what you would have felt to be satisfactory? ☐

   (c) below what you would have felt to be satisfactory? ☐

.....

Reactions to ASSP - Post-results COMMUNITY/EMPLOYER

Name of your organisation _____

Date _____

1. Have you read or heard any information about the results of the recent nationwide testing program on reading, writing and numeracy (Australian Studies of Student Performance)?

   ☐ Yes By what means? _____

   ☐ No

2. Has your organisation had discussions on the ASSP program and it's results?

   ☐ Yes On approximately how many occasions? _____

   ☐ No

   Are such discussions scheduled for further meetings?

   ☐ Yes

   ☐ No

3. In the reports that you have seen or heard,

   (a) was the restricted range of the topics tested made clear?

   ☐ Yes

   ☐ No

   (b) were the ways in which the results can and cannot be interpreted made clear?

   ☐ Yes

   ☐ No

4. Does your organisation believe that some sort of nationwide testing program is useful.

   ☐ Yes What sort? _____

   Why? _____

   ☐ No Why not? _____

   ☐ No official opinion

5. Has your organisation made any approaches to any of the following with a view to discussing the ASSP results?

   (a) schools ☐ Yes ☐ No

   (b) state education officials ☐ Yes ☐ No

   (c) government members ☐ Yes ☐ No

   (d) parliamentary opposition members. ☐ Yes ☐ No

   Does your organisation intend or plan to make any such approaches in the near future?

   Yes ☐ No ☐ Not decided ☐

6. How does your organisation feel about the results of the ASSP testing?

   For the 10 year olds, in overall terms were the results:

   (a) satisfactory? ☐

   (b) better than satisfactory? ☐

   (c) below what would have been satisfactory? ☐

   For the 14 year olds, in overall terms were the results:

   (a) satisfactory? ☐

   (b) better than satisfactory? ☐

   (c) below what would have been satisfactory? ☐

INDEX

Accountability 1-3, 8, 16, 117

Assessment of Performance Unit 2, 108, 119

Australian Capital Territory 64, 71

Australian Co-operative Assessment Project 19, 20, 94, 104, 121

Australian Council for Educational Research 2, 9-11, 14-17, 19-28, 54-58, 66-70, 73-79, 88, 94-97, 118, 119

Australian Council for Educational Standards 11, 42

Australian Education Council 2, 4, 14, 20-23, 27, 37, 70, 94, 95-98, 116

- Standing Committee 22-23
- Steering Committee 23-28, 118
- Technical Sub-Committee, 24-25, 57, 118

Australian Studies in School Performance 9-11, 27, 32

Australian Studies in Student Performance

- Aims 2-3, 10, 14, 23, 37-38, 44, 116-117
- Cost 20, 23, 25, 30, 119
- Design, scope 3, 23-27, 39-43, 57-66, 118-119
- Expectations 17, 29-30, 38, 42-46, 50, 51, 121, 122
- Genesis 1-3, 8-28
- Impact 75, 90-96, 118-121
- Implementation 75-79, 86-89, 119
- Management 23-28, 117
- Participation 64, 72-75, 88
- Political uses 97
- Results 95-96, 120-121
- Technical issues 57-66, 120
- Validity 29, 55, 105-108, 118-119

Australian Teachers Federation 11, 14, 26-28, 37, 97, 105, 106, 120

Basic Skills 1, 3, 9, 10, 12, 17, 53, 55, 96, 116

Control 14, 15, 26, 27

Commonwealth Department of Education 26

Commonwealth Government 7, 10-14, 16-21, 28, 97

Community Groups View 113

Co-operative Federalism 21

Conservativism 1, 2, 8-11, 15

Criterion-referenced testing 3, 10, 19, 28, 39, 55-59, 118-119

Curriculum 2, 16, 26, 30, 45, 90, 92, 99, 103, 107, 122

Curriculum Development Centre 29, 37, 42, 117

Cut-off points 39, 59-61, 88, 94, 119

Difficulty of Items 57-61, 107

Directors-General of Education 17-25, 28, 31, 97, 100

Economic Factors, Role of 8-10, 14, 15, 28, 102

Education Research and Development Committee 3, 9, 12, 15-17, 19, 22, 118, 119, 122

- Study Group Report 15-17

Employers 10, 47, 49

Essential Skills Assessment Project 18, 19

Ethnic Groups 25, 26

Evaluation Study

- Aims 3-5, 116
- Design 37-38, 46, 70-71, 77, 84, 85, 92, 98, 108
- Origins 3, 4

Fitzgerald Report 13

House of Representatives Select Committee 9, 12-13, 16, 53

Learning Difficulties 9, 10, 13

Literacy and Numeracy 8-14, 39, 53-56, 97-101, 123

Mastery 3, 40

Media, Role Played by 11, 27, 98-103, 120

Minimum Competency 1, 39, 52-61, 101

Ministerial Responsibility 8, 25, 30

National Assessment 4, 5, 15, 29, 30, 37, 72

National Assessment of Educational Progress 1, 2, 16, 21, 104, 119

National Estimates 61-66

New South Wales 17, 31, 65

Non-government Systems 26, 37, 49, 50, 101

Northern Territory 27, 31

Numeration 56, 95

Parents, Parents Organizations 11, 37, 45, 47, 80, 81, 92, 111, 119, 121

Performance 59, 60, 106, 107

Politics 8-12, 15, 20, 21, 23, 50, 96-98, 105, 114, 119, 120, 121, 122

Pressure Groups 10

Principals 77, 87-90, 108

Professional Associations 48, 102, 108, 113

Progress and Review Testing 20-24

Progressive Ideology 9-14

Quality of Education 14, 100-101, 123

Questionnaires
- Pre-Testing 71-77
- Implementation 77-83
- Initial Impact 84-94
- National Impact 108-112

Queensland 17, 18, 31, 64, 65, 103

Rasch Model 2, 6, 61, 108

Reading 55, 95

Remediation 13, 121

Response Rates 64, 65

Results, National 95-101

Sampling 25, 61-63

School Achievement Testing Proposal 19-20

Schools Commission 9, 11, 12, 25, 42

Schools, Awareness of ASSP 71
- Impact of ASSP 75, 86-92, 108-120, 121
- Implementation of ASSP 75-83
- Participation in ASSP 70-74

Senate Standing Committee 28

South Australia 65, 99, 103, 104

Standards 1-3, 8-17, 28, 38-40, 97, 98, 100-102, 116, 121

State Testing 17

Student Reactions 81-83

Tasmania 17, 64, 65, 71

Teacher Education 13

Teachers Unions 18, 29-31, 42, 47-48, 64, 73, 105-106, 119

Teachers Views 79, 80, 92, 111

Teaching, Impact on 89-92

Test Development 57-59, 118, 119

Victoria 18-19, 27, 31, 64, 104-106

Western Australia 17, 65, 75, 99

Williams Report 14

Writing 56, 95